Robots and Racism: Examining Racial Bias towards Robots

Arifah Addison

HIT Lab New Zealand

University of Canterbury

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Supervisors

Chrisoph Bartneck (HIT Lab NZ), Kumar Yogeeswaran (Dept of Psychology)

Thesis in partial fulfilment of the requirements for the degree of

Master of Human Interface Technology

June 2019

Abstract

Previous studies indicate that using the 'shooter bias' paradigm, people demonstrate a similar racial bias toward robots racialised as Black over robots racialised as White as they do toward humans of similar skin tones (Bartneck *et al.*, 2018). However, such an effect could be argued to be the result of social priming. The question can also be raised of how people might respond to robots that are in the middle of the colour spectrum (i.e., brown) and whether such effects are moderated by the perceived anthropomorphism of the robots. Two experiments were conducted to examine whether shooter bias tendencies shown towards robots is driven by social priming, and whether diversification of robot colour and level of anthropomorphism influenced shooter bias. The results suggest that shooter bias is not influenced by social priming, and interestingly, introducing a new colour of robot removed shooter bias tendencies entirely. Contrary to expectations, the three types of robot were not perceived by the participants as having different levels of anthropomorphism. However, there were consistent differences across the three types in terms of participants' response times.

Publication Co-Authorship

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form - Masters

Chapter/section/pages of this thesis that are extracted from co-authored work and details of the publication or submission from the extract comes:

The results of the experiments in this thesis were presented at the Artificial Intelligence, Ethics and Society (AIES) Conference in Honolulu, Hawaii in January 2019. As of May 2019 it is currently in the process of being published as part of the conference proceedings under the title "Robots Can Be More Than Black And White: Examining Racial Bias Towards Robots" (screen shot provided).

The chapter/sections below were included in the paper, and expanded on in the final thesis:

- Abstract
- Introduction
- Chapter 4: Experiment A – Social Priming and Fatigue
- Chapter 5: Experiment B – Diversification and Anthropomorphism
- Chapter 6: Conclusions, Limitations and Future Work

Nature and extent of contribution by the candidate: **90%**

The research presented in this thesis is a progression of a prior study which was published by the thesis supervisors, Christoph Bartneck and Kumar Yogeeswaran (and others) as "Robots and Racism". Subsequently they had outlined some further studies, which the candidate was able to pick up and carry out for their MHIT project. The co-supervisor Kumar Yogeeswaran was included due to both being a co-author on the initial study, and being from the psychology department and therefore able to provide the required guidance in regard to carrying out a psychological study.

The experiments were variations of those used in the prior study, with the initial ideas for the new ones expanded on and developed further by the student and supervisors together. The student was responsible for the development and carrying out of all the experiments. The methodology structure was based on the previous study, with improvements and enhancements to the supporting processes able to be made by the candidate due to their existing skills in programming and data transformation (from a previous career in database development and data analysis). The supervisors initially drove the final statistical analysis and interpretation of the results, with the student learning from them such that they could perform it subsequently.

Therefore, all steps were either carried out by the student after discussion and agreement with the supervisors, or done in collaboration, with the student acquiring new skills. The supervisors contributed to the results write-up where applicable, to ensure the accuracy and appropriateness of the interpretation of the results in terms of both a psychological study and a Human Interface Technology paper. See the table provided detailing the relative contributions in POS format.

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all.

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Masters candidate's contribution to this co-authored work.
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text.

Name: **Christoph Bartneck** Signature: (On behalf of both supervisors) Date:

Name: **Kumar Yogeeswaran** Signature: Date: 27/6/19

See Appendix A for confirmation of publication and contribution details.

Contents

List of Figures

List of Tables

Chapter 1: Introduction and Literature Review

Hugo and Nebula-winning science fiction author David Gerrold foresees a world where Artificial Intelligence (AI) is part of every-day life, where a robot teddy bear may be your child's ever evolving playmate and teacher (Gerrold, 2018). Artificial intelligence agents and social robots are indeed becoming more of a reality, with future use predicted in areas such as education, therapy, and help for the physically impaired (Weir, 2018). These are roles in which there is a need to relate to and connect with the client, which raises questions around how these robotic assistants should be developed. Is it important for them to not merely be likeable and friendly but to project specific traits? Should they demonstrate features of a particular gender, personality or culture in order to help achieve their goals? How will they be perceived by the clients, and could it be in ways that were not considered by those that built them? To answer these questions, the developers will need to work alongside social scientists and psychologists (Weir, 2018).

Some research has been done around how the gender and personality of artificial intelligence is perceived and reacted to, but there has been little work on the area of race. This is likely because most robots developed so far are not very human-like. However, those that are could be seen to have a race, as shown in Figure 1.1, and the first studies have begun to emerge on the topic. For example, by analysing free-form comments posted about highly human-like robots in online videos, Strait *et al.* (2018) found that people dehumanised robots racialised as Asian and Black more frequently than they did robots racialised as White. Gong (2008) found that in an Implicit Association Test (IAT) using computer-generated images of humans and robots, White participants who showed a racial prejudice against the Black human images also tended to prefer robots over Black humans. This was contrary to their initial theory that robots would be seen as an out-group to humans by all participants. But as the robot stimuli were almost all silver in tone, could the participants have been subconsciously perceiving them as White?

Figure 1.1: ERICA - ERATO ISHIGURO Symbiotic Human-Robot Interaction Project

Racial bias involves a complex relationship between people's conscious overt beliefs, and their subconscious covert attitudes. This means that simply asking them whether they are biased is not getting a true picture, so tools have been developed to measure implicit, or subconscious bias that the participant may not even realise they have. One of these tools is the Shooter Bias.

The Shooter Bias paradigm was first introduced by Correll *et al.* (2002) in a paper titled *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals.* Correll was inspired by cases of police in the USA mistakenly shooting innocent people, and the increasingly prevalent idea that these mistakes may be race-based, even if on a subconscious level. If the suspect is more closely associated with crime and violence in the minds of the officers they are more likely to be identified as a potential threat. If the police are more fearful of African American males, they may be more likely to think they are holding a weapon, and therefore quicker to open fire on them. This theory was tested in an laboratory experiment using a simple computer game that measured the participants' implicit, or subconscious responses. When they were asked to 'shoot' images of Black or White human males that were holding a gun, and 'not shoot' those holding benign objects (e.g. a wallet or cell phone), a clear bias was found. Participants of all races showed a tendency to be quicker to shoot armed Black men than White, and quicker to not shoot unarmed White men than Black. The effect was collectively referred to as a Shooter Bias, via an interaction between the target's race (Black or White) and the object they are holding (gun or benign object). A recent meta-analysis of 42 studies in the psychological literature (Mekawi & Bresin, 2015) supports this general tendency of a shooter bias.

Researchers in human computer interaction wondered if the shooter bias was carried over to humanoid robots, and showed that this was indeed the case in the paper "Robots and Racism" (Bartneck *et al.*, 2018). The experiment was carried out using both the human images from Correll's original 2002 studies, and 'racialised' robots that were recoloured with Black and White human skin tones, and the results were found to be comparable to the original human-only study. Specifically, people showed a similar shooter bias toward robots racialised as Black relative to White in a similar fashion as they showed toward Black vs. White humans, no matter their own race or explicit racial views.

The following sections review literature around racial bias where it pertains to this topic - i.e unconscious or covert bias, and more recently emerging studies on how people subconsciously perceive social robots and artificial intelligence.

1.1 Racial Bias

Racial bias and how it is expressed falls along a spectrum from blatant and overt to subtle and covert, and people also adapt to what is acceptable to be shown in their society as the norms change over time (Yogeeswaran *et al.*, 2017). This can manifest as either not voicing strong views that may not be well received, or generally being careful about how they talk about race so as not to offend. They often also have subconscious biases they aren't even aware of at all.

The term 'Aversive Racism' was coined by Samuel L. Gaertner to describe people who consciously regard themselves as non-prejudiced but at the same time possess unconscious negative feelings and beliefs about minority groups (Gaertner & Dovidio, 2005). These attitudes can manifest in subtle and insidious ways. When it is obvious that a particular choice or action would be biased they will do the 'right thing', but if they can rationalise a negative response in some other non-race based way they will do so. For example, they found that when White people thought they were the only witness to a staged emergency they helped both Black and White victims equally, but when there were other witnesses around who could also get involved they helped White victims twice as much as Black.

I.e., they were more likely to leave the black victims for other people to help.

1.2 Perception of Social Robots

Tay *et al.* (2014) investigated how stereotypes of gender and personality affected acceptance of social robots in roles that are typically female vs male dominated - health-care and security. In each of the two roles, robots displayed combinations of gender (male vs female) and personality type (introvert vs extrovert) via non-verbal cues. The results showed that gender, personality and their respective role stereotypes all interacted to affect user acceptance. Participants initially preferred robots with gender and personalities that matched the stereotype of each role, but over the course of the interaction, personality had a stronger effect than gender. For example, the female health-care agent was initially more trusted than the male. However, if the female also seemed introverted, by the end of the interaction it was less trusted than the male extrovert. Conversely the male security agent was more trusted over the female unless it expressed as introverted.

Eyssel & Kuchenbrandt (2011) found that people rated social robots more favourably and also more strongly anthropomorphised those that belonged to their in-group, in comparison to those that were in the participant's out-group. In another study, researchers investigated the use of a fully autonomous social robot as a one-on-one learning companion for preschool children. They measured the childrens valence (positive/negative reaction) and engagement using facial expression analysis. The data was fed directly to the robot so it could learn and adapt accordingly to personalise the motivational strategies it used for each child. They found that this significantly increased valence compared to interactions with a non-personalising robot (Gordon *et al.*, 2016).

Tapus *et al.* (2007) defined socially assistive robotics as those that help human users primarily through social rather than physical interaction. They predicted their use in the assistance of the elderly and physically impaired, in rehabilitation, and for people with cognitive disabilities, developmental or social disorders. They focused on six areas of the agent embodiment, personality, empathy, engagement, adaptation and transference. Ultimately, the users should be able to apply what they learn from their interactions with the agent to their interactions with other people. The researchers ask "what models from psychology, cognitive science, and social science can be effectively utilised to help the goals of social assistive robotics?".

People with autism have difficulty learning appropriate social behaviour and early intervention can make a significant difference to their lives. However, it is difficult to detect before the age of three, and therapies must be very particular and consistent in order to be effective. Social robots can be programmed to be very precise and perfectly repeatable, which is exactly what autistic children need. They can monitor eye contact with greater accuracy and for longer time than a human, and send the data to be analysed. They can use play exercises that help children improve their social, sensory and cognitive skills which are then transferable to real-world situations (Cabibihan *et al.*, 2013).

These findings, when combined with those on implicit bias, help bring together the idea that we need to consider how we design social robots with regard to how they are perceived in terms of race. The current research explores this idea by asking some key questions as outlined in the following section.

1.3 Current Research Questions

The current research expands on the previous work by Bartneck *et al.* (2018) by addressing some issues in the original robot racism experiments, and also answers some new questions posed by the researchers. Was the bias shown towards the darker robots actually an effect of social priming? Would a range of colours along the racial spectrum have different results? Would the bias vary depending on how human-like the robots are? Another factor that may have influenced the robot experiment was the longer game length. Adding the robot images as well as the original human ones doubled the amount of trials the participants had to do, which may have caused player fatigue.

Social Priming

Would such a shooter bias effect emerge if humans were removed from the task entirely? Ogunyale & Howard (2018) argued that the robot shooter bias experiment was a "classic case of social priming" because participants completed ratings of Black Americans and White Americans before completing the shooter bias task. Additionally, participants made judgements about shooting both Black and White humans and robots in the same task, so stereotypes about Black people may have simply been applied to robots of the same skin tone. Therefore, one of the primary goals of the present work was to address this criticism by having participants complete the shooter bias task prior to any questions about race or ethnicity and by removing the humans from the experiment entirely so that participants complete the task with the robot images only.

Colour Diversity

Does shooter bias follow a continuum from dark to light skin tones, or is it largely based on stereotypes and prejudices that people have toward a certain group? Previous research suggests that there is a strong shooter bias towards Black males even when Latino and Asian males are present as well as Whites, but there was no significant difference found between Latino, Asian and White males (Sadler *et al.*, 2012). Therefore, the present work examined whether shooter bias would be particularly evident toward robots racialised as Black relative to those that are Brown or White.

Anthropomorphism

Do shooter bias effects observed by Bartneck *et al.* (2018) depend on the degree of perceived anthropomorphism of the robot? Anthropomorphism has an important role in HRI. As shown by Zlotowski & Bartneck (2013), people cognitively process robot faces as they do human faces. Therefore might people demonstrate stronger shooter bias tendencies toward robots that are more human-like relative to those that are more machine-like? As the previous work had only used one robot type in the experiments (i.e., the Nao robot), it is unclear whether shooter bias effect would vary as a function of the perceived anthropomorphism of the robot (i.e., from less to more human-like). The present work addresses this question by examining shooter bias using three different robots (i.e., Inmoov, Nao, Robosapien) that the researchers perceived as varying on a continuum of anthropomorphism from human-like to machine-like.

Fatigue

Previous work has shown that the more cognitively fatigued people are, the more their implicit biases play a part in their decisions, and Macrae *et al.* (1994) suggested that stereotyping evolved to help preserve mental processing resources. Gilbert & Hixon (1991) took Caucasian participants who were native English speakers and asked them to perform a word-fragment completion test. Some words in the test were stereotypically associated with Asian people, and they were helped by either Asian or Caucasian assistants during the test. By keeping some of the subjects cognitively busy by having to rehearse an eight-digit number, they showed that cognitive busyness both inhibits activation of stereotyping and increases the likelihood that any activated stereotypes will be applied.

Correll *et al.* (2013) found that cognitive fatigue has an influence on shooter bias, but they investigated pre-game cognitive loading rather than game length. Therefore, the number of trials as a measure of fatigue can be examined to see whether it has an influence on shooter bias. This also serves to reduce the risk of unknown variables across experiments.

Chapter 2: Background

The current research is a progression of the prior “Robots and Racism” study by Bartneck *et al.* (2018) outlined in the introduction. The researchers took the well-established social-psychological paradigm known as the Shooter Bias and ran their own version of the experiment to see if the bias was transferred to similarly coloured robots.

2.1 Original Shooter Bias Experiment

The original human-only Shooter Bias study by Correll *et al.* (2002) used a simple computer game to test people’s reactions to potential threats. When playing the game, each subject was shown a series of images of Black or White males against a variety of backgrounds holding either a gun or a benign object such as a wallet or cell phone (see Figure 2.1, images from Millisecond.com). They were instructed to press a key on the keyboard corresponding to whether they chose to ‘shoot’ or ‘not shoot’ each image. They had less than a second to make their decision, and their response in terms of accuracy (was it a correct response) and latency (how long did they take to hit the key) for each trial was recorded.

Figure 2.1: Examples of game stimuli: Black or White men holding a gun or benign object

Results

As reported in the study, reaction times for all correct trials were log-transformed and averaged for each subject. A 2×2 analysis of variance (ANOVA) of the trials with Ethnicity (Black vs. White) and Object type (Gun vs. Benign Object) as within-subject factors showed a significant main effect for Object. Subjects were both significantly faster to correctly respond to targets holding a gun than they were to those with a benign object. There was also a significant main effect for Object type x target Ethnicity interaction. Participants of all races shot at armed Black targets faster than they did armed White targets, and chose to not shoot unarmed White targets faster than unarmed Black targets.

Further Experiments and Overall Conclusions

The original shooter bias study included three further experiments, the first was a repeat of the initial experiment but using a shorter response window (650ms) and stronger incentives (loss or gain of money based on accuracy) and the error rates were analysed for accuracy. The results showed that people more often mistakenly shot an unarmed Black target, and more often mistakenly refrained from shooting an armed White target.

In order to see whether there was a relationship between the participants' explicit views and the shooter bias, they repeated the initial experiment with the addition of a survey on racial and social attitudes. The survey included questions regarding their explicit prejudice against African Americans, and their willingness to express any prejudices they held. They were also asked how much they viewed African Americans to be aggressive or dangerous. A between-subjects factor of Participant Ethnicity (African American vs. White) was also added to the analysis, to see if the race of the subject affected the shooter bias. The bias was shown towards Black men by both Black and White participants.

When the experiment was run with additional Asian and Latino targets Sadler *et al.* (2012), the shooter bias was still strongly shown towards the Black males but not towards the other three ethnic groups. Over all their studies, the researchers found the shooter bias towards African American males was present for US participants no matter their motivations, racial attitudes or own race.

2.2 Initial Robots and Racism Study

Bartneck *et al.* (2018) noticed that most robots on the commercial market seemed to be mostly white or metallic in colour. They wondered why this was the case, and whether the white robot could resemble a race. I.e., would they be perceived as 'White'. Furthermore, would they show the same implicit racial bias towards darker coloured robots as they do towards Black men in the shooter bias paradigm?

Experiments and Results

The first two shooter bias experiments from the Bartneck *et al.* (2018) original human study were replicated, but with the additional targets of humanoid robots that had been 'racialised' by re-colouring them with the skin tones of African and Caucasian women from a professional photograph of multiracial women (Bartneck *et al.*, 2018, Figure 2). Therefore, participants were now shooting at both human and robot targets, as per the examples shown in figure 2.2.

Figure 2.2: Original human targets and new 'racialised' robots

They also surveyed the participants on their explicit views on race and ethnicity, and gathered their personal demographics including ethnicity, age, gender, religion and nationality. All participants were recruited from the USA via Crowdflower (now FigureEight), and they were offered incentives as per the initial experiment (a baseline payment for participation, plus prizes for the best players). In addition, they performed a manipulation check of whether subjects would explicitly attribute race to a robot. They were shown each colour of robot and asked to choose which race it was from a list, the first option being 'does not apply'.

Using the same analysis as in the original human study, it was found that subjects responded to the 'racialised' robots in a similar way to their human counterparts.

Racialisation: 2x2 ANOVA of Race (Black vs White) x Object in Hand (armed vs unarmed) showed a significant 2-way interaction. Post-hoc t-tests showed that participants were quicker to shoot armed Black agents than armed White agents, and slower to refrain from shooting unarmed White agents than unarmed Black agents

Agent Type: 2x2 ANOVA of Agent (human vs robot) x Object in Hand (armed vs unarmed) showed a significant 2-way interaction. Post-hoc t-tests showed that participants were quicker to shoot armed Humans than armed Robots and slower to refrain from shooting unarmed Humans than unarmed Robots.

Explicit Bias: Participants showed overall more positive attitudes toward White Americans than Black Americans although they did not show a significant difference in their own personal stereotypes of seeing either race as aggressive or dangerous. However, they thought that others would - i.e., they thought that Blacks would be seen as more of a threat by others than they did themselves.

Robot Race: Most participants did choose to ascribe a race to the robots, and did say that the darker robot was Black (70%) and the lighter one White (67%). Only 11% chose 'does not apply'.

2.3 Current Research

Reviews of the initial Robots and Racism study raised several questions:

1. Did having the participants shoot at both human and robot targets mean that social priming was a factor in the transference of the bias?
2. Did the difference in the length of the experiment resulting from the addition of the robot images mean that player fatigue may have affected the bias?
3. Would the addition of robots representing other races affect the bias, as this was not shown to be the case in human studies.
4. Additionally, does the level of anthropomorphism of the robot affect shooter bias?

Chapter 3: Methodology

Correll's original experiment used 80 participants in the USA, and subsequent studies showed that the shooter bias towards black men is only present in the USA. For the robot shooter bias studies, online platforms were used to recruit subjects and run the experiments, which allowed access and restriction to US participants. This enabled the final running of the experiments to be extremely fast compared to running them in a laboratory setting, but also meant that a lot of time and care had to be taken in testing and refining the procedure.

Each experiment consisted of a demographics survey, the shooter bias game, and post-game manipulation checks (ascription of race or rating the perceived level of anthropomorphism of the robots). This chapter describes the tools used to build and carry out the experiments and data analysis, and how they fit together into the overall methodology of the project. The majority of the time on the project was spent developing and refining the procedure and the actual running of each experiment was done over a few hours.

The main phases involved were to build the experiment, create the online recruitment project, run the experiment, and analyse the data. The overall project process is shown in figure 3.1.

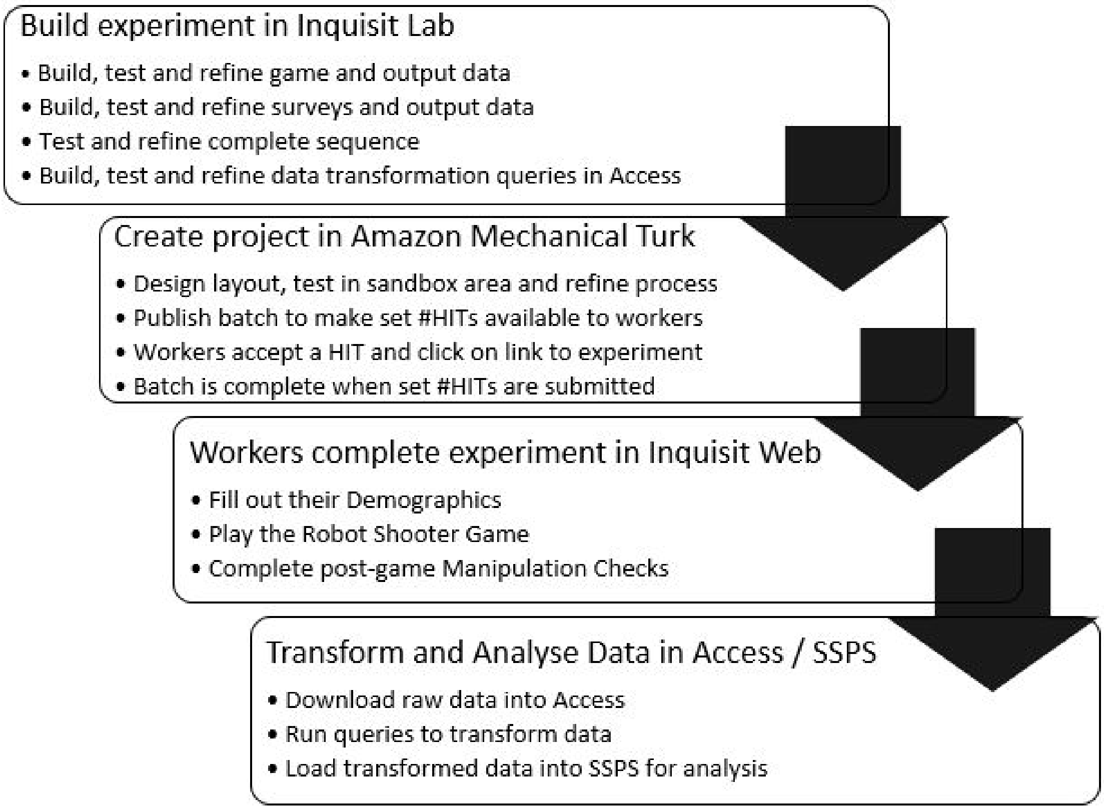


Figure 3.1: Project process overview

Each of these phases is outlined in more detail in the following section, using examples from the experiments as applicable.

3.1 Building the Experiments with Millisecond Inquisit

Millisecond Inquisit is software that was developed for researchers to carry out cognitive, social, neuro-physiological, and psychological experiments. Participant response data is recorded to the millisecond which makes it perfect for running the shooter bias experiments. It consists of Inquisit Lab to develop and run experiments on a local computer, and Inquisit Web to run them online. Millisecond provides a library of experiments or 'tests' to be used or adapted as required, which includes Correll's original shooter bias experiment, the "Police Officer's Dilemma" (Figure 3.2).

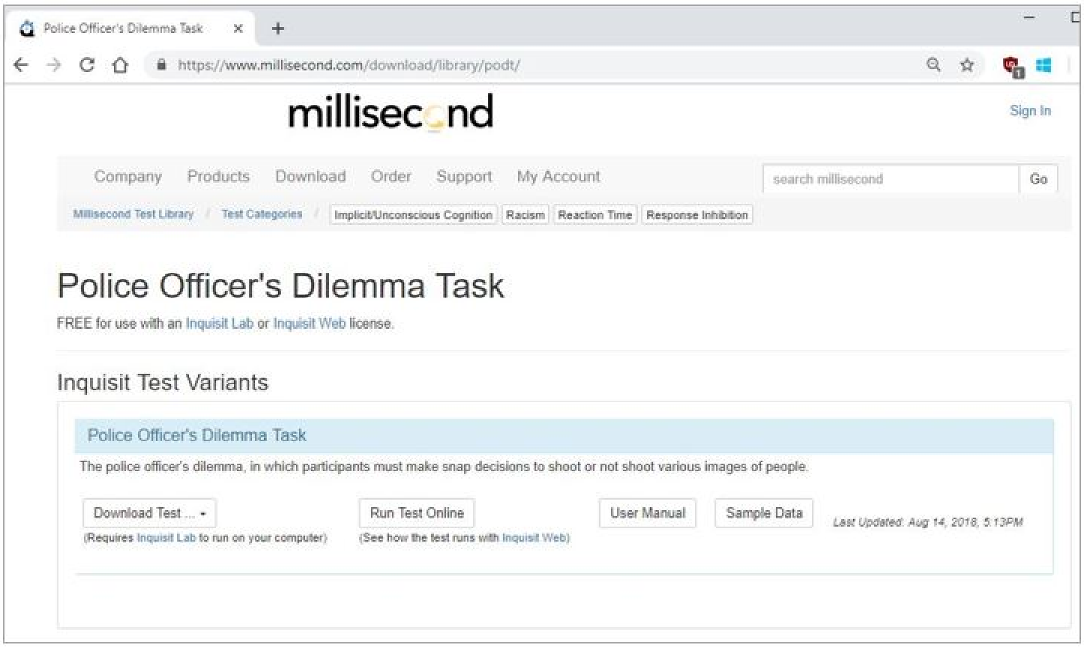


Figure 3.2: Millisecond test library - The Police Officer's Dilemma task

The Police Officer's Dilemma was downloaded from the test library for the initial Robots and Racism experiment, which included the test script and image files. It was then adapted by adding in the robot images and altering the code and data output parameters accordingly. This test was then adapted further for the current experiments. The development process for each new test took many hours of testing and refinement as every change had to be carefully checked to ensure there were no ripple effects causing undesired consequences. Once completed, the experiment's files (test script and images) were ready to be loaded into Inquisit Web (see Appendix B).

The groups specified in the script's code had to correspond with the Subject ID settings for the task. For example, Experiment B has three groups for the between-subjects analysis (one for each robot type), with two sets of each group to counterbalance the shoot/not shoot keys for left/right hand (i.e., half the participants used the left-hand 'A' key to shoot the targets and the right-hand 'L' key to 'not shoot', and vice-versa). Therefore there are six groups in total which must be specifically set in Inquisit (Appendix B.8).

3.2 Recruiting Participants: Amazon Mechanical Turk

Amazon Mechanical Turk (MTurk) is an online 'crowdsourcing marketplace' which people (known as 'workers') sign up for to get paid to complete tasks created by 'requesters'. Mturk is often used by academic researchers to recruit participants for online surveys and experiments. A 'sandbox' area is provided where projects can be tested and refined before publishing on the live site. There are three main phases of creating a projects - setting the properties, designing the layout, and publishing it as as a HIT (Human Intelligence Task). Mturk workers can then view and accept the HIT if they wish to participate.

In order to run the experiments for this study, a project called "Robot Shooter Game" was created. Properties that were set up (Figure 3.3) included the reward per assignment, i.e., the amount each participant will be paid upon completion (in USD as MTurk is a US-based company), and the 'Number of assignments per task', i.e., the amount of participants required.

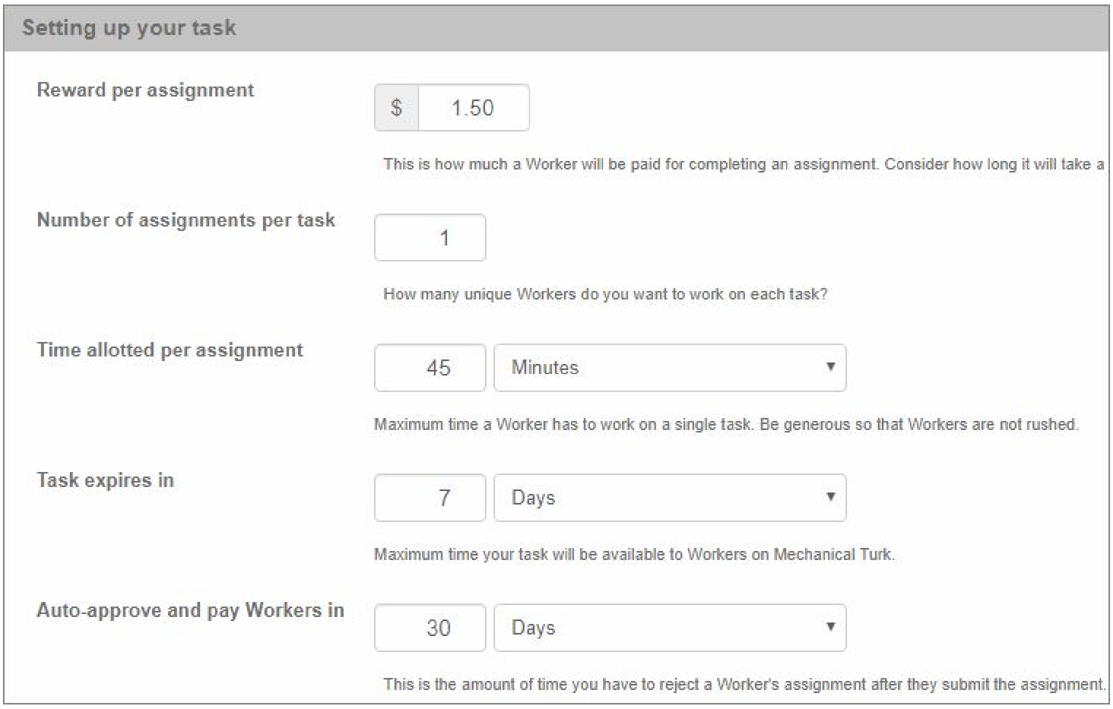


Figure 3.3: MTurk project - set task properties

The amount of time the batch was available to workers, and when they got auto-approved (and therefore paid) were set to ensure the HIT was available long enough to get the full amount of participants, and also give enough time to analyse their response data for suspicious patterns and reject them if necessary (to help maintain worker quality).

Once the Robot Shooter Game project was created, it was able be run as many times as needed in separate 'batches'. Each batch could have different property settings, so there was no need to create a new project for each variation of the experiment. For example, the first batch was a pilot run with a small number of participants (10), then any necessary adjustments could be made before running a second batch for the rest - with the number of participants setting altered to make up the total required.

Worker Requirements were then set (Figure3.4), which included qualifications that restricted the availability of the experiments to only desired workers. They had to be in the USA in order to replicate the original shooter bias experiments, and also have a minimum 99% approval rate and at least 50 HITs approved to help ensure a high standard of worker. In order to prevent workers taking part in more than one Robot Shooter Game experiment, they also had to have not been granted the 'Played_Game' qualification. This was allocated to every worker who became a participant, therefore excluded them from seeing any subsequent Robot Shooter Game HITs.

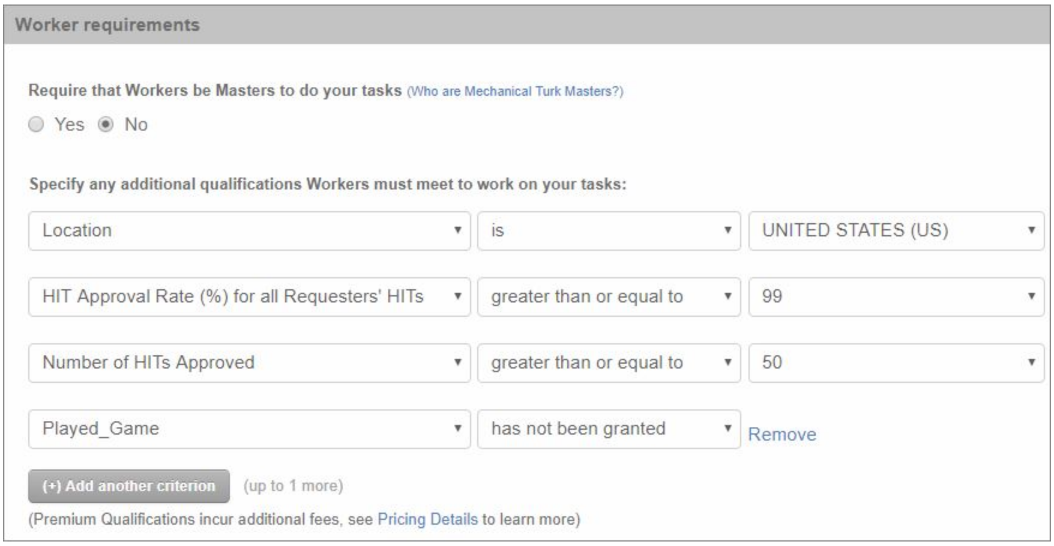


Figure 3.4: MTurk project - set worker requirements

The page for the HIT that is displayed to workers was then designed (Figure 3.5).

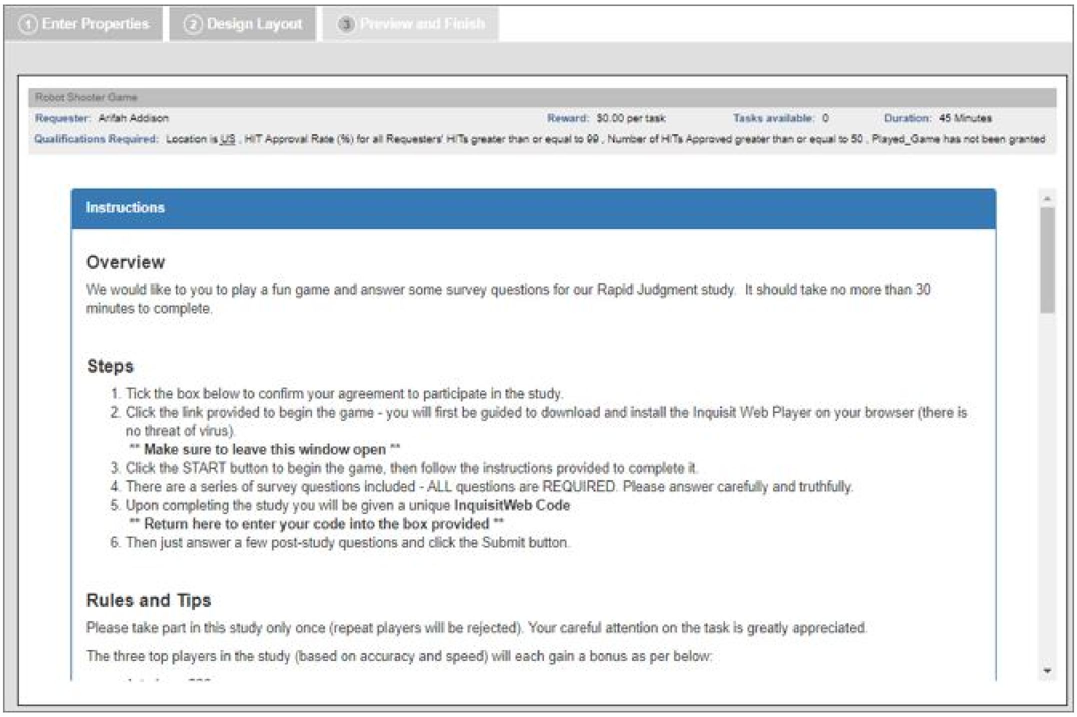


Figure 3.5: Mturk project - preview page layout

Once finalised, the batch details were displayed. In the example shown in Figure 3.6, no assignments have yet been completed so the progress bar is at 0%.

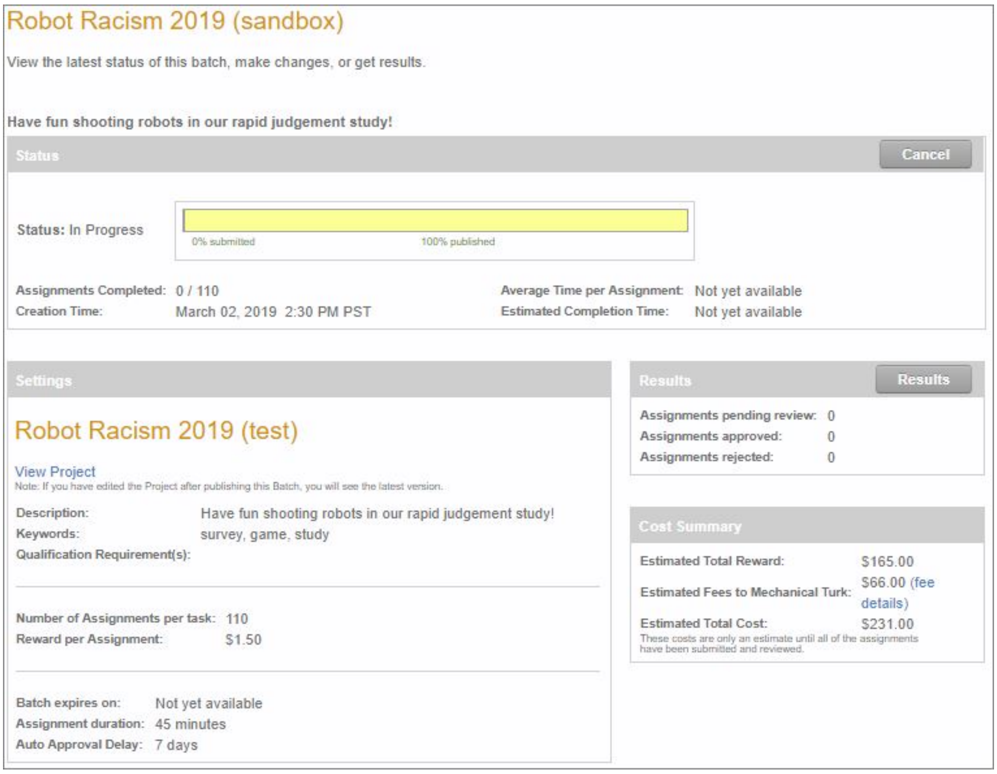


Figure 3.6: Mturk project - published batch details

Qualified MTurk workers were then able to see the Robot Shooter Game HIT and click 'Accept & Work' to take part in the experiment (Figure 3.7).

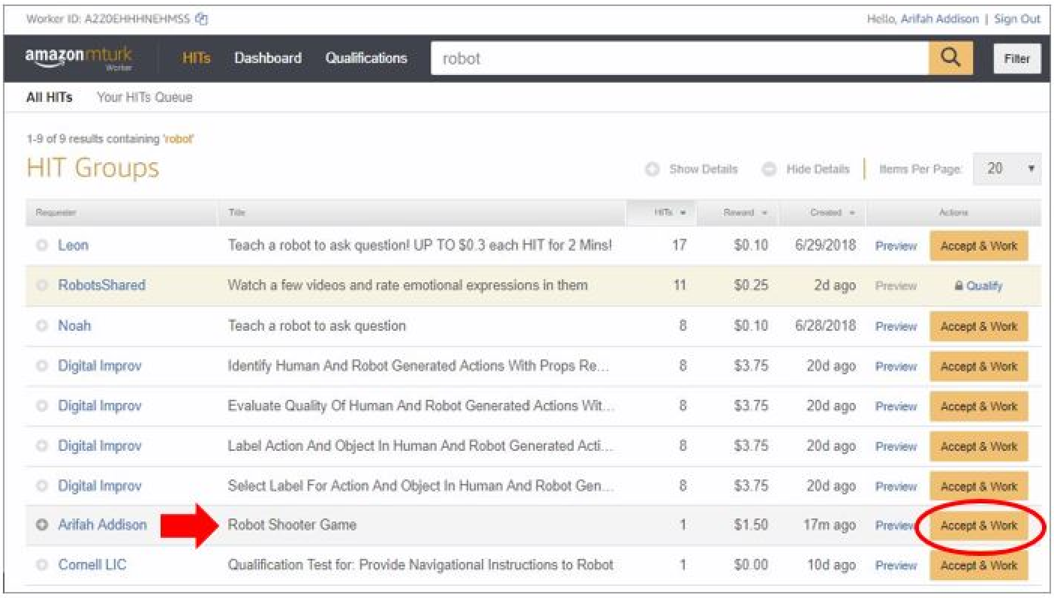


Figure 3.7: Worker HIT List with Robot Shooter Game

The following section shows the workflow process from a participant's perspective.

3.3 Completing the Experiment - Participant Workflow

Only qualified workers as per the settings in Figure 3.4 were able to see a Robot Shooter Game HIT, most importantly ensuring that they are located in the USA and that they do not take part in more than one of the experiments.

3.3.1 Overview

There are three main phases involved for the workers to complete the experiment:

1. Accept the HIT in MTurk and click the link to run the experiment.
2. Complete the experiment via InquisitWeb - including filling in demographics, playing the robot shooter game and completing manipulation checks.
3. Return to MTurk and submit their results.

These phases and the complete workflow process from a participant's perspective is summarised in figure 3.8 below.

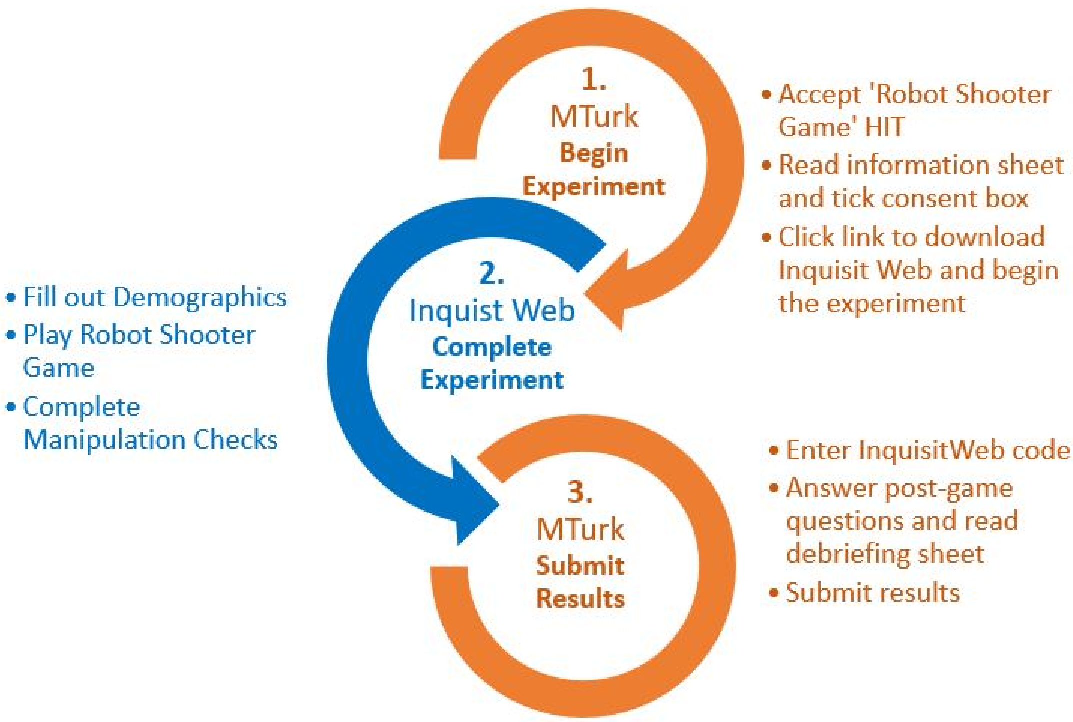
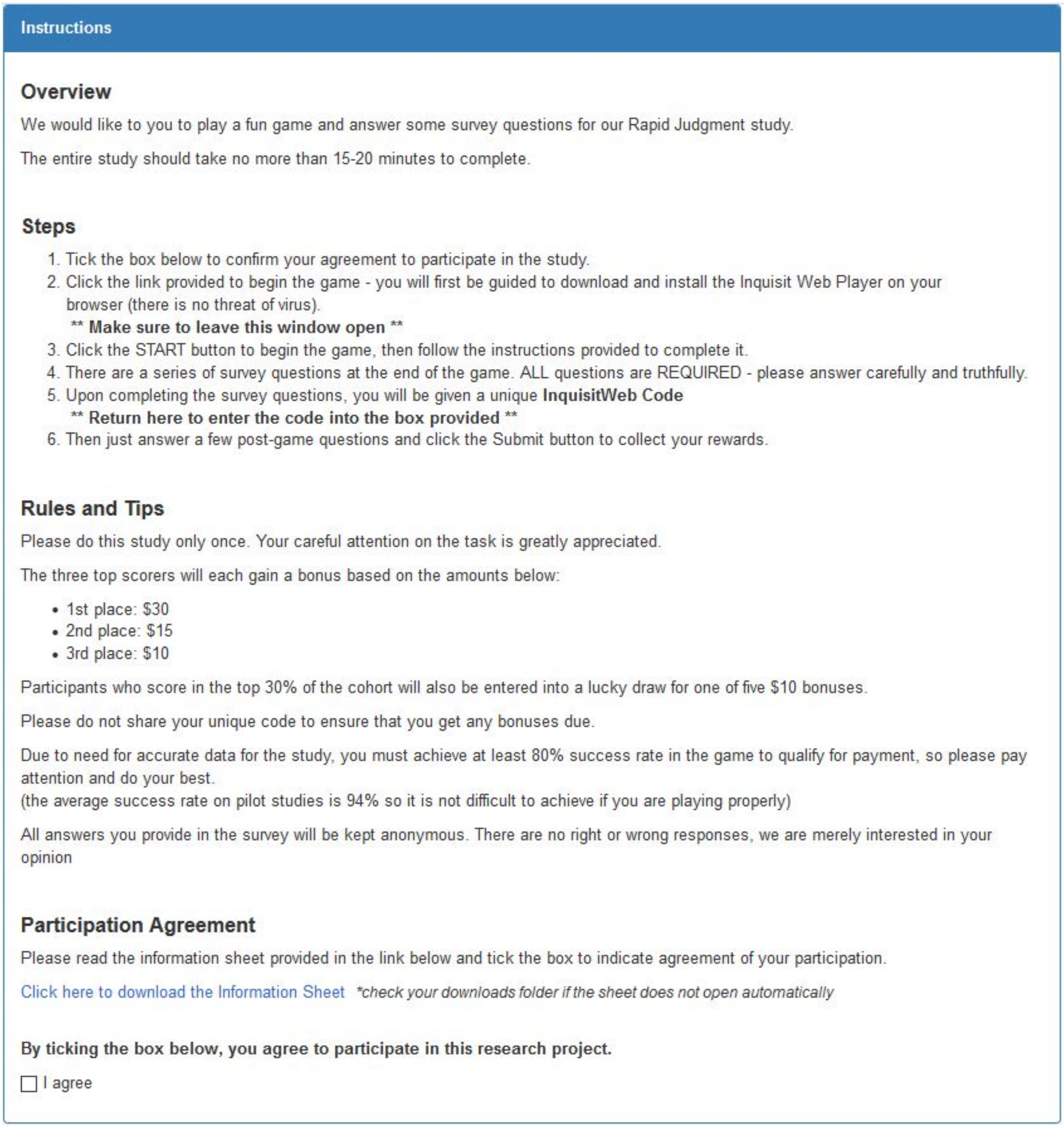


Figure 3.8: Participant workflow

3.3.2 Phase 1 - Begin Experiment in MTurk

After finding the Robot Shooter Game HIT in Mechanical Turk, qualified workers were able to review and accept it. The task page displays general instructions which include rules and tips to encourage serious participation in the game and a list of potential prizes (Figure 3.9). In order to get official consent from the workers they had to tick a box to confirm their agreement to participate. A link to the information sheet is available.



Instructions

Overview

We would like to you to play a fun game and answer some survey questions for our Rapid Judgment study.

The entire study should take no more than 15-20 minutes to complete.

Steps

1. Tick the box below to confirm your agreement to participate in the study.
2. Click the link provided to begin the game - you will first be guided to download and install the Inquisit Web Player on your browser (there is no threat of virus).
   **** Make sure to leave this window open ****
3. Click the START button to begin the game, then follow the instructions provided to complete it.
4. There are a series of survey questions at the end of the game. ALL questions are REQUIRED - please answer carefully and truthfully.
5. Upon completing the survey questions, you will be given a unique **InquisitWeb Code**
   **** Return here to enter the code into the box provided ****
6. Then just answer a few post-game questions and click the Submit button to collect your rewards.

Rules and Tips

Please do this study only once. Your careful attention on the task is greatly appreciated.

The three top scorers will each gain a bonus based on the amounts below:

- 1st place: $30
- 2nd place: $15
- 3rd place: $10

Participants who score in the top 30% of the cohort will also be entered into a lucky draw for one of five $10 bonuses.

Please do not share your unique code to ensure that you get any bonuses due.

Due to need for accurate data for the study, you must achieve at least 80% success rate in the game to qualify for payment, so please pay attention and do your best.
(the average success rate on pilot studies is 94% so it is not difficult to achieve if you are playing properly)

All answers you provide in the survey will be kept anonymous. There are no right or wrong responses, we are merely interested in your opinion

Participation Agreement

Please read the information sheet provided in the link below and tick the box to indicate agreement of your participation.

Click here to download the Information Sheet **check your downloads folder if the sheet does not open automatically*

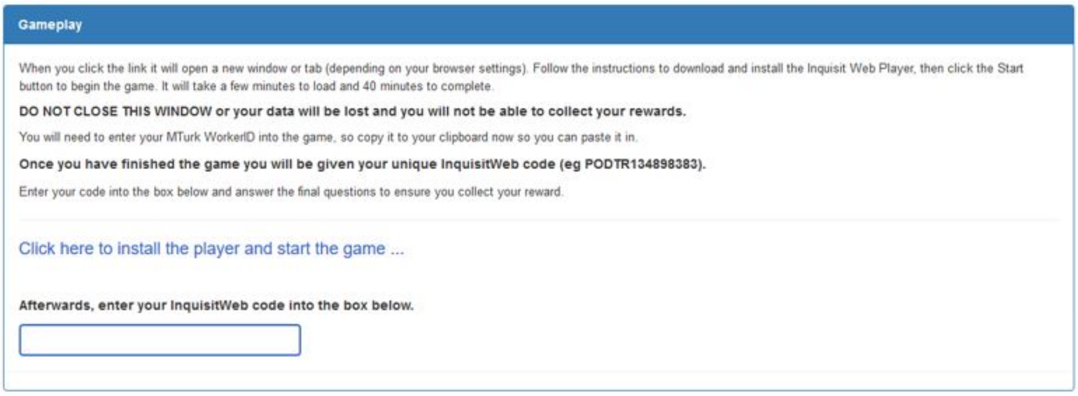
By ticking the box below, you agree to participate in this research project.

☐ I agree

Figure 3.9: General instructions

Gameplay instructions were also displayed (Figure 3.10), including warnings to keep the Mturk window open as they had to return to it to enter their post-game unique InquisitWeb code and submit their results. The InquisitWeb code is so their MTurk data can be cross-referenced with their Inquisit data, and ensure nobody can attempt to claim payment if they did not complete the experiment.

If the agreement box has been ticked, the participant was able to click the link to install the InquisitWeb player and begin the experiment.


Gameplay

When you click the link it will open a new window or tab (depending on your browser settings). Follow the instructions to download and install the Inquisit Web Player, then click the Start button to begin the game. It will take a few minutes to load and 40 minutes to complete.

DO NOT CLOSE THIS WINDOW or your data will be lost and you will not be able to collect your rewards.

You will need to enter your MTurk WorkerID into the game, so copy it to your clipboard now so you can paste it in.

Once you have finished the game you will be given your unique InquisitWeb code (eg PODTR134898383).

Enter your code into the box below and answer the final questions to ensure you collect your reward.

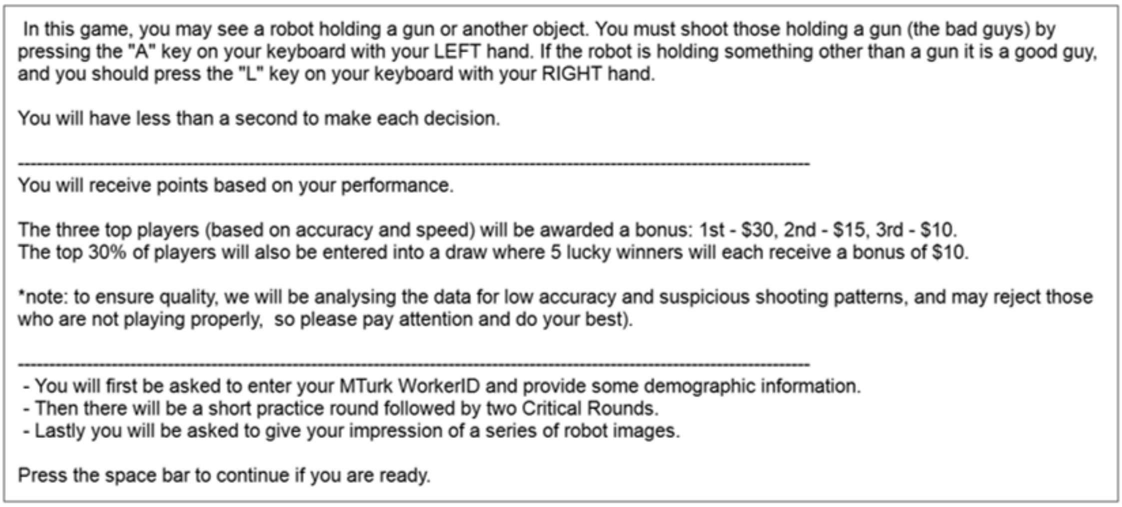
Click here to install the player and start the game ...

Afterwards, enter your InquisitWeb code into the box below.

Figure 3.10: Starting the experiment

3.3.3 Phase 2 - Complete the Experiment in Inquisit

The first screen of the experiment shown in figure 3.11 displayed the specific instructions on how to play the game, and reiterated the prizes and encouragement to play with integrity. It explained that the participant must first enter their demographic information, then they will have a practice round followed by the critical phase (1 or 2 rounds depending on the experiment). Finally they will be required to give their impressions of some robot images (i.e., complete the manipulation checks).


In this game, you may see a robot holding a gun or another object. You must shoot those holding a gun (the bad guys) by pressing the "A" key on your keyboard with your LEFT hand. If the robot is holding something other than a gun it is a good guy, and you should press the "L" key on your keyboard with your RIGHT hand.

You will have less than a second to make each decision.

You will receive points based on your performance.

The three top players (based on accuracy and speed) will be awarded a bonus: 1st - $30, 2nd - $15, 3rd - $10.
The top 30% of players will also be entered into a draw where 5 lucky winners will each receive a bonus of $10.

*note: to ensure quality, we will be analysing the data for low accuracy and suspicious shooting patterns, and may reject those who are not playing properly, so please pay attention and do your best).

- You will first be asked to enter your MTurk WorkerID and provide some demographic information.
- Then there will be a short practice round followed by two Critical Rounds.
- Lastly you will be asked to give your impression of a series of robot images.

Press the space bar to continue if you are ready.

Figure 3.11: Game instructions

Before playing the Robot Shooter Game, participants had to first enter their demographic details such as gender, age, and race, and also their MTurk WorkerID so their Inquisit data can be cross-referenced with their MTurk data. This also meant that if a participant emailed saying they couldn't submit their results on the MTurk side (e.g., if they closed the window accidentally), they could still be paid if they gave us their correct Inquisit code.

First they had a short practice round to get familiar with the gameplay before beginning the critical rounds, of which there are either one or two depending on the experiment. Examples of robot stimuli and backgrounds they may have encountered (depending on the experiment and group they are in) are shown in figure 3.12 below.

Figure 3.12: Examples of robot stimuli

For each robot they encounter, they had to press the appropriate key to either ‘shoot’ if they think it has a gun, or ‘not shoot’ if they think it is carrying something else.

As per the original Correll experiment, participants got one of the following messages if they correctly chose to shoot an armed robot (Good Shot!) or to not shoot an unarmed robot (Wise Choice!).

Good Shot!
You win 10 points

Wise Choice!
You win 5 points

They got one of the following messages if they incorrectly chose to shoot an unarmed robot (You shot a good guy!) or to not shoot an armed robot (You’re Dead!).

You shot a good guy!
You lose 20 points

You're Dead!!!
You lose 40 points

If they didn’t hit either of the designated keys in the 850ms window, they got a message saying they are Too Slow.

Too Slow
You lose 10 points

The running total of their current score was also displayed, and between each phase they had appropriate messages or instructions to ensure a smooth flow (see Appendix B).

After participants finished the robot shooter game they completed the manipulation checks. For both experiments this included ascribing a race to each colour of robot encountered in the game (e.g figure 3.13).

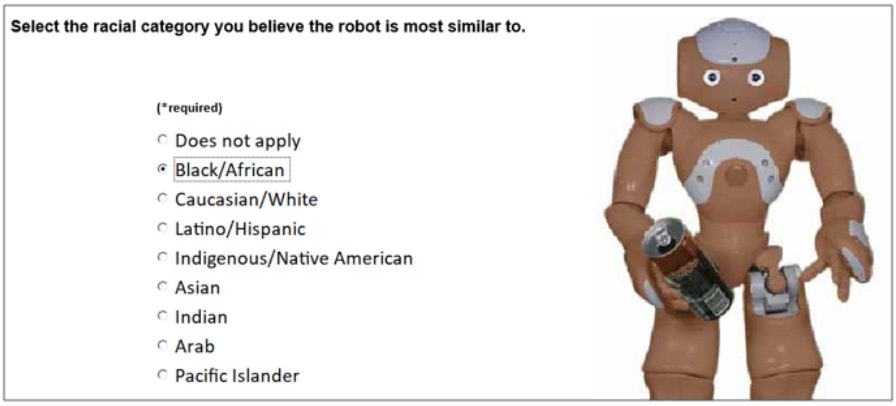


Figure 3.13: Manipulation check - Ascription of Race

For Experiment B they were also asked to evaluate how human-like they thought each type of robot to be using the anthropomorphism questions from the Godspeed Questionnaire (e.g. figure 3.14). The questionnaire was developed by Bartneck *et al.* (2009) as a tool to measure people's perception of anthropomorphism and other traits of robots. They evaluated a neutral (true white) robot of each type.

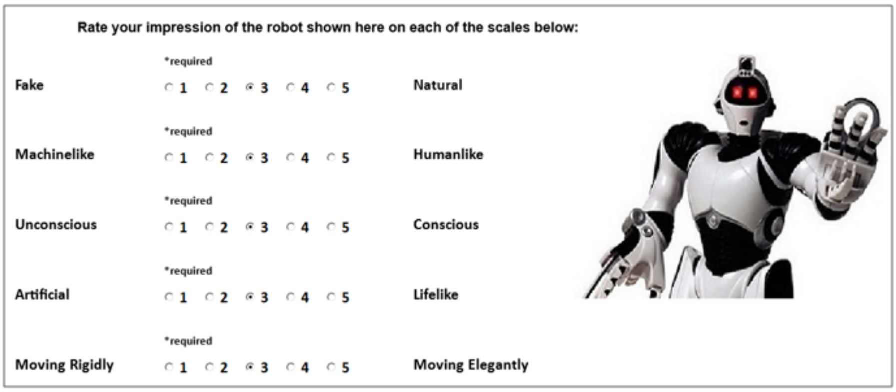


Figure 3.14: Manipulation check - Anthropomorphism

For each check, participants were also asked to rate each robot in terms of their perceived danger and trustworthiness.

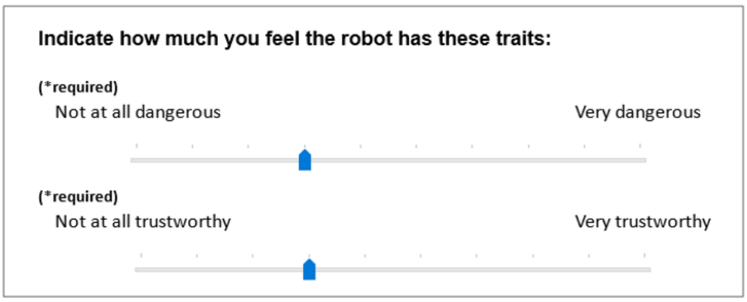
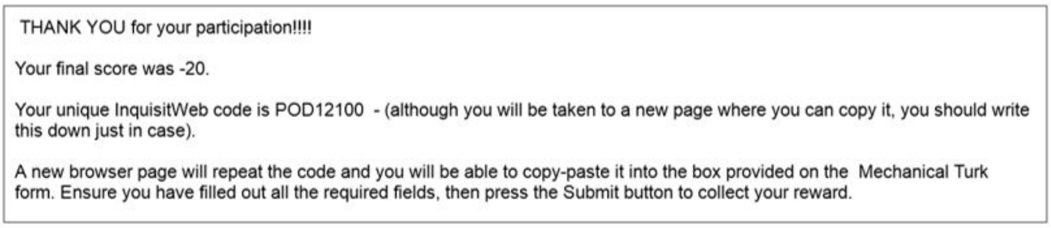


Figure 3.15: Manipulation check - Traits

Once the experiment was completed the participant was given their unique InquisitWeb code (Figure 3.16), generated using a formula based on their SubjectID and Group Number.


THANK YOU for your participation!!!!

Your final score was -20.

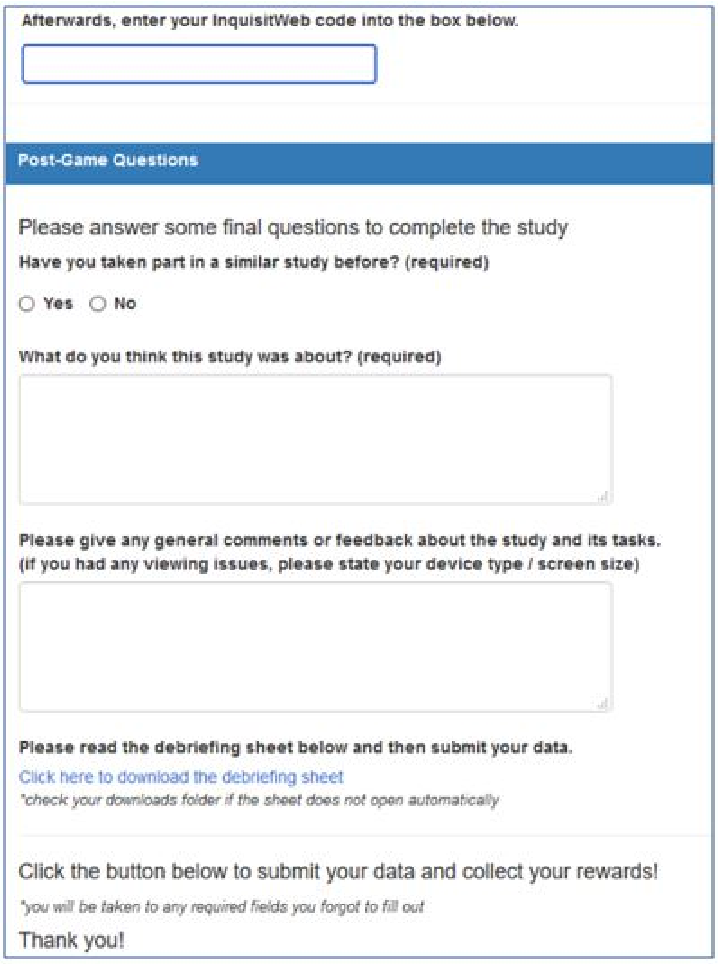
Your unique InquisitWeb code is POD12100 - (although you will be taken to a new page where you can copy it, you should write this down just in case).

A new browser page will repeat the code and you will be able to copy-paste it into the box provided on the Mechanical Turk form. Ensure you have filled out all the required fields, then press the Submit button to collect your reward.

Figure 3.16: End of experiment in Inquisit

3.3.4 Phase 3 - Submit Results in MTurk

The participant then returned to the Mechanical Turk page to enter their Inquisit code, and they had the option of filling in some post-game questions and reading the debriefing sheet (Appendix A.3) before submitting their results (Figure 3.17).


Afterwards, enter your InquisitWeb code into the box below.

Post-Game Questions

Please answer some final questions to complete the study

Have you taken part in a similar study before? (required)

Yes No

What do you think this study was about? (required)

Please give any general comments or feedback about the study and its tasks.
(If you had any viewing issues, please state your device type / screen size)

Please read the debriefing sheet below and then submit your data.

Click here to download the debriefing sheet

*check your downloads folder if the sheet does not open automatically

Click the button below to submit your data and collect your rewards!

*you will be taken to any required fields you forgot to fill out

Thank you!

Figure 3.17: Enter unique Inquisit code into Mturk page and submit results

Once the full amount of required participants submitted their results, the data was available for downloading and analysis, as outlined in the next section.

3.4 Analysing the Data

When an experiment is run in Inquisit Web, the data from each participant is automatically collected and available for download. The main steps of data analysis after an experiment was completed are:

- Download the data from Inquisit Web as CSV (comma-separated values) files.
- Import the CSV files into a pre-prepared Microsoft Access Database. Also import the MTurk data for the corresponding HIT for cross-referencing.
- Clean the data - run queries to check for any anomalies (e.g., cross-check with the MTurk data and check for WorkerIDs that do not have a match).
- Run queries to transform the cleaned data into the format required for statistical analysis and output into an Excel spreadsheet.
- Import the Excel spreadsheet data into SSPS Statistics for statistical analysis.

The overall process is summarised in figure 3.18.

Figure 3.18: Data analysis process

After completing the experiment, the data was available for download from the Inquisit Web site. There was initially a separate file for every participant, each containing many rows for each of the experiment trials. These were then downloaded as single files for the first stage of the analysis in MS Access. There were files for each of the raw and summary data outputs from Inquisit, and the MTurk Batch file for cross-referencing the WorkerID. This file also contained their answers to the post-game survey questions on the Mturk side of the process, and the unique Inquisit code they entered in order to submit their results. Each data file contained output fields as specified in the experiment design to record the participant responses and associated data, as well as general system information (time and date etc). See Appendix C for a full data dictionary of the raw and summary files.

Cleaning and Transforming the Data in Microsoft Access

Microsoft Access is a database management system designed for the development of small scale professional database applications, but in this project it was used for the first stage of data analysis. Queries were developed to integrate and transform the raw data into the required format for statistical analysis.

First the data files were imported into tables in the database. This data was then cleaned - e.g., fixing incorrect spellings of their country or race so these fields can be grouped. There were also queries designed to look for anomalies - e.g., MTurk Worker

IDs in the Inquisit data that did not exist in the MTurk batch data, or double-ups of the MTurkID (meaning they had done it twice), or InquisitWeb code (if they had given their code to another worker). The cross-referencing also enabled identification of workers who completed the experiment in Inquisit, but accidentally closed their MTurk window and could not submit their results. It could be confirmed that they had played the game if they provided their Inquisit code and MTurkID, and a separate HIT was created to in order to pay them.

The initial raw data was in ‘long format’; i.e., there is one row for each participant response, so there are many rows for each subject. For example, there are 93 rows for each participant who fully completed Experiment B, including their survey answers and gameplay responses. As this is not in a configuration that can be analysed by SPSS Statistics, it must be converted into ‘wide format’ with one row per participant.

As well as having to aggregate the many rows of data for each participant into one, it must also be grouped by block, as the values in each field may have different meanings between them. For example, the ‘trialcode’ and ‘response’ fields for the Demographics block contains the data they entered for the survey, but the same fields for a game-play block will show which stimulus condition they were shown and key they pressed in response to it. Figure 3.19 shows an example of the Demographics block and the first few rows from the first gameplay block for a particular participant.

Some of the fields are only relevant for the game-play, so either have no value or can be excluded from analysis of survey data. For example, the ‘correct’ and ‘latency’ fields are irrelevant for the demographics questions although a value is recorded. The number of gameplay trials they got correct is used to calculate the mean Success Rate (SR) as a measure of Accuracy, and the latency in milliseconds between when the stimulus was shown and when they hit the key is used to calculate the mean Response Time (RT).

blockcode	b	trialcode	tr	va	vi	v	val	v	response	corr	value	latency	valu
Demographics	1	mturkid	1	0	1	0	0	0	A3S5ZB4D9ξ	1		9330	0
Demographics	1	gender	1	0	1	0	0	0	Female	1		9330	0
Demographics	1	age	1	0	1	0	0	0	29	1		9330	0
Demographics	1	race	1	0	1	0	0	0	White/Cauc	1		9330	0
Demographics	1	raceother	1	0	1	0	0	0		1		9330	0
Demographics	1	nationality	2	0	1	0	0	0	American	1		12316	0
Demographics	1	religion	2	0	1	0	0	0	None	1		12316	0
Demographics	1	state	2	0	1	0	0	0	Arizona	1		12316	0
Demographics	1	hand	2	0	1	0	0	0	Right	1		12316	0
POD_1	3	ObjectWhiteWoowee	7	2	2	4	800	4	30	1	CR	443	5
POD_1	3	ObjectBlackWoowee	12	4	4	2	1000	2	30	1	CR	503	10
POD_1	3	GunBlackWoowee	16	3	3	1	1000	1	38	1	Hit	450	20
POD_1	3	GunWhiteWoowee	20	2	2	1	600	1	38	1	Hit	390	30

Figure 3.19: Example of raw data trial responses

The game-play data was averaged for each condition per participant, to calculate their Success Rates (correct responses) as a measure of Accuracy and their mean Reaction Times for each experimental condition (colour of robot x object in hand). These are the measures that are used for the final analysis. Therefore, each block's data must be extracted separately, then they all must be combined into a final output with data to be one row per participant.

Due to the nature of the raw data, designing the queries to transform it was fairly complicated. Figure 3.20 shows the Access window, with the data tables and queries in the left panel, and the final output query design in the main area. There are seven queries that go into this final one, and some of them will also be made up of other queries themselves. However, once fully tested, the database was simply duplicated with empty tables for each experiment. So after gathering the data, it was loaded in to a fresh copy of the database and instantly transformed into the final output format.

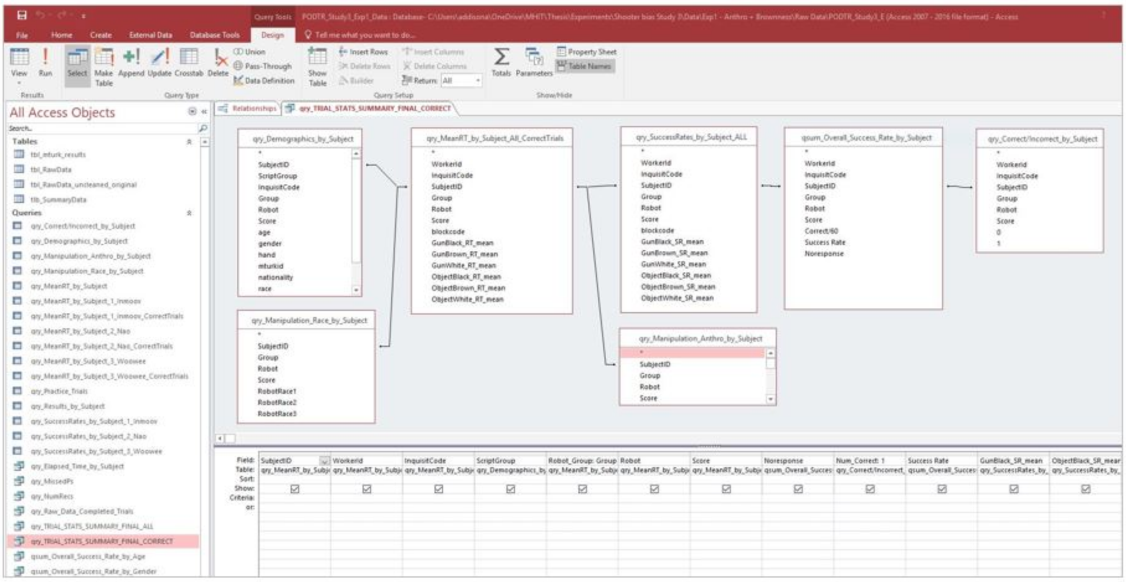


Figure 3.20: Contents of Access database with final query design

Figure 3.21 shows some of the output of this query, with one row for each participant.

qry_TRIAL_STATS_SUMMARY_FINAL_CORRECT

SubjectID	WorkerId	InquisitCode	Scri	Robc	Robot	Score	Noresponse	Num_Correct	Success Rate	GunBlack_SR_mean	ObjectBlack_SR_mean	GunBrown_SR_mean
607246421	A2C25LZB76\	POD31821741360	3	2	Nao	-515	34	13	21.67%	0.30		0.50
948588870	A3774HPOUI	POD12845768707	1	1	Inmoov	-655	29	17	28.33%	0.20	0.50	0.20
973337017	A322Y1IBX2L	POD32920013148	3	2	Nao	-370	33	19	31.67%	0.50	0.10	0.40
611324861	A2IOYG5PYLI	POD31833976680	3	2	Nao	-390	24	22	36.67%	0.70	0.10	0.50
798510054	A4LIJVRU6DG	POD32395532259	3	2	Nao	-745	1	27	45.00%	0.80	0.30	0.20
845833745	A2USG66URL	POD62537503332	6	3	Wowee	-495	4	29	48.33%	0.80	0.40	0.40
132097435	A9IC7BMHIP	POD1396294402	1	1	Inmoov	-585	2	29	48.33%	0.60	0.40	0.70
814743805	A2F0A4FIVT9	POD52444233512	5	3	Wowee	-250	18	31	51.67%	0.70	0.70	0.60
101418088	A3QEVFM3UI	POD1304256361	1	1	Inmoov	-550	7	31	51.67%	0.60	0.70	0.40
69897040	AA74LC8YEG!	POD1209693217	1	1	Inmoov	-285	8	33	55.00%	0.50	0.60	0.90
740561728	A31CYXT7WH	POD32221687281	3	2	Nao	-340	1	33	55.00%	0.70	0.30	0.80
725897472	A3TZNS3TG0	POD22177694513	2	1	Inmoov	-55	20	35	58.33%	0.70	0.60	0.50
678409289	A1GFN7WDJ(	POD42035229964	4	2	Nao	-270	1	35	58.33%	1.00	0.40	0.80
907047243	A1ROIF2ZSIU	POD32721143826	3	2	Nao	-300		37	61.67%	0.80	0.60	0.70
638644526	ASXFMHGSFC	POD51915935675	5	3	Wowee	-390	1	37	61.67%	0.30	0.70	0.90
525960402	A1GVHFB9F2	POD31577883303	3	2	Nao	-395	1	38	63.33%	0.60	0.80	0.60
868601890	ARYEH1BMQ	POD12605807767	1	1	Inmoov	-350		38	63.33%	0.80	0.50	0.50
405192045	A53S7J4JGW	POD21215578232	2	1	Inmoov	-150	5	41	68.33%	0.40	0.90	0.60

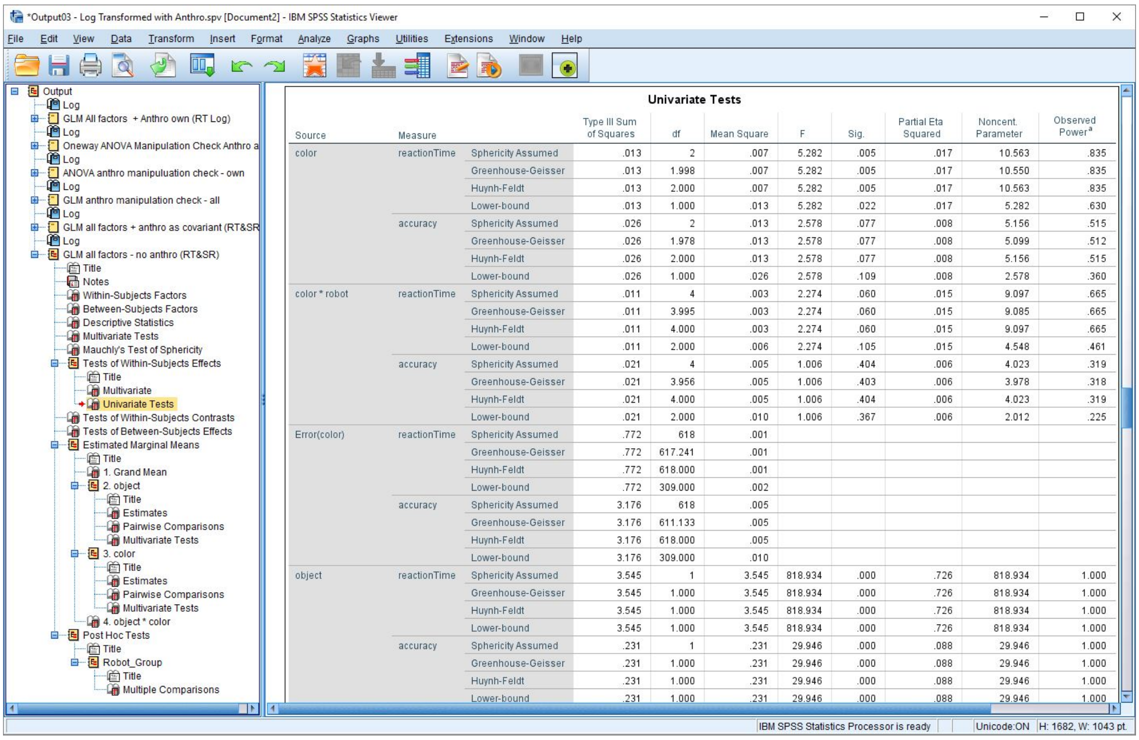
Figure 3.21: Snapshot of final query output data

As per the original human and previous Robot Racism experiments, only data for participants who successfully completed all phases of the experiment is included, and only correct trials analysed. The final dataset is exported into an Excel spreadsheet to be prepared for importing into SPSS Statistics for analysis. This preparation involves inspecting the data for quality responses. For example, some participants had a high no-response rate (i.e., not hitting any keys) or showed a suspicious pattern (e.g., hitting the same key repeatedly). After sorting by their mean success rate and final score, and inspecting their responses, it was decided that only including those with 80% or higher in the analysis would effectively filter out all those who had not been playing well enough to be good test subjects[1].

Final Analysis

IBM SPSS Statistics was used to perform the statistical analysis on the transformed data, in order to show whether the experiments had significant results or not. The final dataset in Excel is imported into SPSS Statistics.

As per the original Correll study and the previous Robot Racism experiments, a General Linear Model or Anova was used (all assumptions being true). In Experiment B there were two within-subjects variables of Object (gun or benign object) and Colour (black, brown and white), with the two measures of Reaction Time and Accuracy. There is also a between-subjects variable of robot type (robot_group 1, 2 or 3). The statistical analysis is then output as per figure 3.22. The various iterations made when inspecting and analysing Experiment B can be seen in the left hand panel.



Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
color	reactionTime	Sphericity Assumed	.013	2	.007	5.282	.005	.017	10.563	.835
		Greenhouse-Geisser	.013	1.998	.007	5.282	.005	.017	10.550	.835
		Huynh-Feldt	.013	2.000	.007	5.282	.005	.017	10.563	.835
		Lower-bound	.013	1.000	.013	5.282	.022	.017	5.282	.630
	accuracy	Sphericity Assumed	.026	2	.013	2.578	.077	.008	5.156	.515
		Greenhouse-Geisser	.026	1.978	.013	2.578	.077	.008	5.099	.512
		Huynh-Feldt	.026	2.000	.013	2.578	.077	.008	5.156	.515
		Lower-bound	.026	1.000	.026	2.578	.109	.008	2.578	.360
color * robot	reactionTime	Sphericity Assumed	.011	4	.003	2.274	.060	.015	9.097	.665
		Greenhouse-Geisser	.011	3.995	.003	2.274	.060	.015	9.085	.665
		Huynh-Feldt	.011	4.000	.003	2.274	.060	.015	9.097	.665
		Lower-bound	.011	2.000	.006	2.274	.105	.015	4.548	.461
	accuracy	Sphericity Assumed	.021	4	.005	1.006	.404	.006	4.023	.319
		Greenhouse-Geisser	.021	3.956	.005	1.006	.403	.006	3.978	.318
		Huynh-Feldt	.021	4.000	.005	1.006	.404	.006	4.023	.319
		Lower-bound	.021	2.000	.010	1.006	.367	.006	2.012	.225
Error(color)	reactionTime	Sphericity Assumed	.772	618	.001					
		Greenhouse-Geisser	.772	617.241	.001					
		Huynh-Feldt	.772	618.000	.001					
		Lower-bound	.772	309.000	.002					
	accuracy	Sphericity Assumed	3.176	618	.005					
		Greenhouse-Geisser	3.176	611.133	.005					
		Huynh-Feldt	3.176	618.000	.005					
		Lower-bound	3.176	309.000	.010					
object	reactionTime	Sphericity Assumed	3.545	1	3.545	818.934	.000	.726	818.934	1.000
		Greenhouse-Geisser	3.545	1.000	3.545	818.934	.000	.726	818.934	1.000
		Huynh-Feldt	3.545	1.000	3.545	818.934	.000	.726	818.934	1.000
		Lower-bound	3.545	1.000	3.545	818.934	.000	.726	818.934	1.000
	accuracy	Sphericity Assumed	.231	1	.231	29.946	.000	.088	29.946	1.000
		Greenhouse-Geisser	.231	1.000	.231	29.946	.000	.088	29.946	1.000
		Huynh-Feldt	.231	1.000	.231	29.946	.000	.088	29.946	1.000
		Lower-bound	.231	1.000	.231	29.946	.000	.088	29.946	1.000

Figure 3.22: Final data analysis output showing Univariate Tests

The Univariate Tests are used to check for significant interaction effects, and as shown in the results sections of the experiments, the interaction between Object and Colour or Robot type are used to check for a Shooter Bias towards them.

[1]Note that they still got paid except for one who had 'no response' for all the gameplay trials.

3.5 Summary of Overall Project Process

For each experiment, every phase of the project was built, tested and refined extensively, then put together into a final implementation process. The general steps taken are shown in Figure 3.1. Although the MTurk sandbox site was used for testing the complete process, pilot experiments were also run on the live site with a small number of real participants as a final check that everything ran smoothly. To this end, the carrying out of the final live experiments and analysis of the data was actually the fastest part of the project.

The MTurk HITs displays instructions and a link to the experiment in Inquisit Web. It also contains an information sheet (Appendix A.1) which refers to the experiment as a 'rapid judgement study'. As is usual in psychological experiments, we are initially misleading the participants as to the real nature of the study so as not to prime them. Applicable workers preview and accept the HIT and do the following:

1. Click on the link provided to load the experiment in Inquisit Web.
2. Fill in demographics, play the game, and complete manipulation checks.
3. Return to MTurk to submit their results.

Before submitting their results, they enter a unique code assigned to them in the game so their MTurk profile can be cross-referenced back to their experiment data in Inquisit. They also fill out some final questions to check whether they have done this kind of study before, what they thought it was really about, and give any general feedback. They can also read a debriefing sheet telling them that the study was about robots and racism, so they can decide not to submit their data if they disagree with it (although that would mean they will not get paid).

Once the specified number of workers have completed the HIT, the experiment is complete and the data is ready for inspection and cleanup before the data analysis process.

In order to quickly clean and transform the data, a Microsoft Access database was prepared and ready. It contained queries that had been developed and refined during the testing and pilot phases of the project.

The overall data analysis process went as follows:

1. Raw data from both MTurk and Inquisit is downloaded into spreadsheet files and imported into Access.
2. Access queries are used to transform and filter the data into the required format and output into a spreadsheet file.
3. Transformed data is loaded into SSPS for final analysis.

Microsoft Excel spreadsheets were used to store both the initial raw data for loading into Access, and the transformed data for loading into SPSS. It was also used to create charts and tables to display the outputs in visual form.

Chapter 4: Experiment A. Social Priming and Fatigue

This experiment sought to find out whether, as suggested by Ogunyale & Howard (2018), social priming may have caused the shooter bias effect in the initial robot shooter bias study. This was achieved removing the possible priming influences - the survey on racial attitudes and stereotypes, and the human stimuli in the game. It also sought to find out whether task length as a measure of fatigue had an influence on the shooter bias effect by analysing the trials in each half of the game independently.

4.1 Method

The initial robot shooter bias experiment was repeated without the human images - i.e., with only the robots racialised as Black and White, and each image displayed in random order within each of 2 blocks of 64 trials (so each image was seen twice, once in each block). The experiment factors were 2 (racialisation: black vs white) $\times$ 2 (object in hand: gun vs. benign object) $\times$ 2 (fatigue: block1 vs block2) within-subjects design. Prior to playing the game, participants provided their basic demographics data, and after the game they were asked to ascribe a race to images of the Black and White robots.

4.2 Participants

A total of 113 participants from the USA were recruited from Amazon Mechanical Turk (MTurk). They received $1.50 USD for completing the experiment and additional bonuses were offered to the top players (based on speed and accuracy). They were also advised that payment may be restricted for those who did not put genuine effort into playing the game correctly (e.g., high no-response rate or hitting the same key repeatedly). After analysing the data accordingly, all those who achieved less than 80% success rate were excluded from analysis, which effectively filtered out all suspicious response patterns. This left a sample of 106 participants (42 female, 64 male). Participant age range was 18 to 58 years ($m = 38; sd = 8.26$). A majority of these participants reported being of White/Caucasian descent ($n = 84$), with others identifying as Black/African American ($n = 9$), Asian ($n = 6$), Latino/Hispanic ($n = 5$), Native American ($n = 2$), and mixed race ($n = 2$).

4.3 Stimuli

The 'racialised' Nao robot images from the initial robot racism study were used. Each robot was either holding a gun, a remote control, a candy bar or a soda can. Using 16 backgrounds $\times$ 2 skin colours (Black vs White) $\times$ 2 objects (gun vs benign object) gave us a total of 64 images.

4.4 Procedure

As outlined in the methodology section, participants were first provided with an information sheet to read and asked to give consent for their participation in the study before being directed to Inquisit Web to complete the experiment.

Participants first answered a series of demographics questions, then went through 20 practice trials with random conditions to ensure a lack of routine. Each trial started with a 500ms fixation, followed by 1-4 empty backgrounds for a random duration (500 to 1000ms), then finally the target image was shown for 850ms. Participants received feedback on how they performed on the trial for 2000ms.

After the practice round, they were allowed to rest and continue to the main test when they were ready - two rounds of 64 trials, each image repeated once per block in random order. After playing the main test the participants were congratulated and given their final score. Participants were then shown images of the Nao robot with each of the two skin tones used in the game in random order and asked to ascribe them a race from a list, including the option of "does not apply" at the very top of the list.

4.5 Measures

The experiment phases included participant demographics, the shooter bias game, and a manipulation check of whether they ascribed a race to each colour of robot.

Demographics

Participants completed a demographic questionnaire which gathered such information as their age, race, gender, nationality, and religion. They also entered their MTurk WorkerID here so it could be cross-referenced back to the recruitment data.

Shooter Bias

Participants' reaction times in deciding whether to shoot/not shoot, and accuracy (correct identification of aggressors versus non-aggressors) while completing the shooter bias task were measured. The reaction time was recorded as the time between the end of the stimuli being shown on the screen and the time when a key was pressed. Following the procedures outlined in Correll *et al.* (2002), the average reaction time and accuracy for different conditions was calculated for the variables of robot racialisation (Black vs. White), and object in hand (gun vs. benign object).

Robot Race

Participants were shown the Black and White Nao robots and asked to ascribe a race to each from a list of options, which included "Does not apply" at the top of the list.

4.6 Results and Discussion

As per the initial robot racism study which followed the procedure in Correll *et al.* (2002), the average log-transformed reaction times for correct trials and the average accuracy rates were analysed for the variables of colour (Black vs White) and whether they are holding a gun or benign object (Armed vs. Unarmed). The two blocks of 64 trials were also analysed separately as a measure of fatigue. The means and standard deviations for all conditions are shown in Table 4.1.

Table 4.1: Means and standard deviations for Reaction Times and Accuracy within Blocks

		Reaction Times		Accuracy	
		Black	White	Black	White
Armed	Block1	581 (52)	582 (51)	0.95 (0.07)	0.95 (0.06)
	Block2	581 (53)	580 (55)	0.97 (0.05)	0.96 (0.06)
Unarmed	Block1	638 (54)	633 (52)	0.91 (0.09)	0.91 (0.09)
	Block2	627 (54)	619 (54)	0.95 (0.07)	0.94 (0.09)

Demographics

Most of the 106 participants accepted for the final analysis identified as White/Caucasian (79%), with 8% Black/African American and the rest either Asian, Latino/Hispanic, Native American or mixed race. 60% of the participants were male, and 40% female. The average age of participants was fairly consistent across all the measures, with an overall average of 34 years old (min 18, max 58). 89% were right-handed and all stated that they were American in nationality.

Table 4.2: Demographics - Experiment A

Race	#	%	**Age**
White/Caucasian	84	79%	35
Black/African American	8	8%	33
Asian	6	6%	35
Latino/Hispanic	5	5%	31
Native American	1	1%	36
Mixed	2	2%	37

Gender	#	%	**Age**
Female	42	40%	36
Male	64	60%	34

Shooter Bias

A 2×2 analysis of variances (ANOVA) revealed the expected significant 2-way interaction between racialization and object in hand for reaction time ($F(1, 105) = 7.428, p = 0.008, \eta^2 = 0.066$), but not for accuracy ($F(1, 105) = 0.114, p = 0.736, \eta^2 = 0.001$). Similar to the previous work of Bartneck *et al.* (2018), participants took significantly longer to not shoot unarmed robots that were racialized as Black compared to those racialized as White ($p < .001$), but there was no difference in time taken to shoot an armed robot that was racialized as Black or White ($p = .89$).

It was also examined as to whether fatigue emerging from task length would impact on shooter bias tendencies. A $2 \times 2 \times 2$ ANOVA revealed non-significant 3-way interaction effect between racialization, object, and fatigue on both reaction time ($F(1, 105) =< .001, p = 0.992, \eta^2 < 0.001$) or accuracy ($F(1, 105) = 0.285, p = 0.595, \eta^2 = 0.003$). However, there was a significant main effect of fatigue on reaction time, ($F(1, 105) = 10.129, p < 0.001, \eta^2 = 0.088$), such that participants were generally faster in the second block than the first block, but this did not interact with robot colour. See Table 4.1) for full results.

Manipulation Check - Attribution of Robot Race

Participants were asked to ascribe a race to both colours of the Nao robot encountered in the game. Although the first option in the list was "Does not apply", 86% of participants did ascribe a race to the robots. The robot racialised as 'White' was mostly identified as White/Caucasian (67%) with the next highest identification being Asian (10%). The robot racialised as 'Black' was mainly ascribed to be Black/African American (47%) or Latino/Hispanic (23%), and 8% identifying it as Indian. This suggests that our manipulation of the skin colour did serve to alter the perceived race of the robot in the eyes of most participants. Counts and percentages of responses are shown in table 4.3.

Table 4.3: Results of manipulation check for Robot Race - Experiment A

	'Black' Robot		**White' Robot**	
Does not apply	*12*	11%	*17*	16%
Caucasian/White	*3*	3%	*71*	**67%**
Black/African	*50*	**47%**	*1*	1%
Latino/Hispanic	*24*	**23%**	*4*	4%
Indigenous/Native American	*0*	0%	*0*	0%
Asian	*2*	2%	*11*	**10%**
Indian	*9*	**8%**	*0*	0%
Arab	*2*	2%	*2*	2%
Pacific Islander	*4*	4%	*0*	0%

4.7 Summary of Experiment A

To summarise, this experiment sought to investigate whether shooter bias towards robots may be a result of social priming by re-running the initial robot shooter bias study with no human images, and with no survey questions regarding participants attitudes regarding race prior to the shooter bias task. The results showed that the shooter bias was still present towards the darker coloured robot even with no social priming, therefore eliminating it as an influencing factor.

The experiment also sought to find out whether playing a greater number of trials influenced the shooter bias, as it had been previously shown that although cognitive fatigue did have an effect task length had not been factored in to any analysis despite a variation of the number of trials in the different experiments. By analysing the data over two blocks independently, it was found that fatigue from trail length did not have an effect on shooter bias tendencies.

Chapter 5: Experiment B. Diversification and Anthropomorphism

This experiment sought to find out whether adding a wider range of human skin tones influenced shooter bias. Previous research with human targets revealed that participants showed similar levels of shooter bias only toward Black males even when including Latino and Asian targets to the original shooter bias paradigm Sadler *et al.* (2012). Therefore it was surmised that including brown coloured robots would not significantly impact the shooter bias tendencies towards the 'Black' ones. The experiment also investigated whether the perceived anthropomorphism of robots interacted with robot colour and object in hand. I.e., it sought to examine whether shooter bias fell across a spectrum of skin tones from darker to lighter, and if this was especially pronounced when a robot was more human-like relative to less human-like.

5.1 Method

The robot shooter bias experiment was repeated with the addition of robots with a brown skin tone (taken from the same image of multiracial women used previously), to achieve a spectrum of racialisation from Black to White. Furthermore, two more types of robot were added (Inmoov and Robosapien) to achieve a range of anthropomorphism from less to more human-like. It was suspected that Inmoov would be perceived as most human-like, followed by Nao, then Robosapien. The experiment was therefore a 3 (skin tone: Black vs. Brown vs. White) $\times$ 2 (object in hand: gun vs. benign object) $\times$ 3 (robot agent: Inmoov vs. Nao. vs Robosapien) mixed design, with robot agent as the only between-subjects factor. Therefore, each participant would encounter only one type of robot but all three colours in the task.

5.2 Participants

A total of 340 participants from the USA were recruited from Amazon Mechanical Turk, and no person was allowed to take part in both experiments. This was done using MTurk's ability to exclude workers with a particular 'Qualification Type' that was assigned to each worker who took part in either experiment. As per Experiment A, only data for all those who achieved at least an 80% success rate was included for analysis. This left a sample of 312 participants (162 female, 150 male). Participant age range was 18 to 78 years ($m = 29; sd = 10.8$). A majority of these participants reported being of White/Caucasian descent ($n = 226$), with others identifying as Black/African American ($n = 35$), Asian ($n = 27$), Latino/Hispanic ($n = 20$), Native American ($n = 2$), and mixed race ($n = 2$). Participants received $1.25 USD for completing the experiment which was slightly less than Experiment A due to the shorter playing time (1 round of 60 trials).

5.3 Stimuli

The three types of robot were all (Inmoov, Nao, and Robosapien) re-coloured with the skin tones of the African, Indian and Caucasian women from the same multi-racial photograph used previously. These were photographed across 5 backgrounds, so the combination of 3 skin colours $\times$ 2 objects (gun vs non-gun) $\times$ 3 robot type (levels of anthropomorphism) which gave us a total of 90 images (30 images for each of the three types of robot).

5.4 Procedure

The experiment was carried out as per the process outlined in the methodology section. The critical round of the game comprised of one block of 60 trials - 5 backgrounds x 6 conditions, each repeated twice in random order. Each participant saw only one type of robot - Inmoov, Nao or Robosapien. After playing the game and ascribing a race to each of the three colours of robot, participants were also shown each of the three types of robot and asked to rate them using the five anthropomorphism questions from the Godspeed Questionnaire Bartneck *et al.* (2009), a standardised measuring tool.

5.5 Measures

Demographics

As per Experiment A, participants completed a demographic questionnaire including questions about their age, race, gender and nationality.

Shooter Bias

As per Experiment A, the reaction time and accuracy across the different trial conditions was measured for the variables of: agent (robot type: Inmoov vs. Nao vs. Robosapien), racialisation (Black vs. Brown vs. White), and object (gun vs. benign object).

Robot Race

Participants were shown Black, White and Brown Nao robots and were asked to ascribe each one a race from a list of options, which also included 'Does not apply'.

Robot Anthropomorphism

Participants were asked to rate their impression of each of the three robot types using five anthropomorphism questions. Each question used a Likert scale from 1-5 where 0=less; 5=more anthropomorphic (i.e., a high rating is more human-like) to get an overall anthropomorphism rating.

5.6 Results and Discussion

As per Experiment A, the average log-transformed reaction times and the average accuracy rates for correct trials were analysed. This experiment had variables of robot type (Inmoov vs. Nao vs. Robosapien), colour (Black vs. Brown vs. White), and whether they are holding a gun or benign object (Armed vs. Unarmed). The means and standard deviations for all conditions are shown in Table 5.1.

Table 5.1: Means and standard deviations for Reaction Times and Accuracy across all conditions

		Reaction Times			Accuracy		
		Black	Brown	White	Black	Brown	White
Armed	Inmoov	558 (58)	552 (55)	548 (58)	0.93 (0.08)	0.93 (0.08)	0.94 (0.08)
	Nao	561 (51)	559 (51)	562 (50)	0.95 (0.07)	0.97 (0.05)	0.96 (0.07)
	Robosapien	538 (55)	541 (55)	536 (54)	0.96 (0.07)	0.97 (0.06)	0.96 (0.06)
Unarmed	Inmoov	600 (54)	595 (54)	600 (53)	0.93 (0.08)	0.93 (0.08)	0.92 (0.10)
	Nao	612 (51)	613 (56)	615 (49)	0.91 (0.12)	0.92 (0.11)	0.92 (0.09)
	Robosapien	595 (51)	583 (46)	586 (51)	0.93 (0.10)	0.94 (0.10)	0.96 (0.07)

Demographics

Most of the 312 participants accepted for the final analysis identified as White/Caucasian (72%), with 11% Black/African American and the rest either Asian, Latino/Hispanic, Native American or mixed race. 48% of the participants were male, and 52% female. The average age of participants was fairly consistent across all the measures, with an overall average of 34 years old (min 18, max 58). 88% were right-handed and 98% stated that they were American in nationality (all were located in the USA).

Table 5.2: Demographics - Experiment B

Race	**#**	**%**	**Age**
White/Caucasian	226	72%	38
Black/African American	35	11%	34
Asian	27	9%	31
Latino/Hispanic	20	6%	30
Indigenous/Native American	2	1%	31
Mixed	2	1%	24
Gender	**#**	**%**	**Age**
Female	162	52%	36
Male	150	48%	36

Shooter Bias

A 2×3 ANOVA revealed a non-significant effect of robot colour (i.e., racialisation) and object on both reaction time ($F(2, 618) = 2.662, p = 0.073, \eta^2 = 0.008$), and accuracy ($F(2, 618) = 0.481, p = 0.619, \eta^2 = 0.002$). Contrary to the previous studies on robot shooter bias, there was no significant difference in shoot and don't shoot responses for robots that appeared Black or White. As the only major differing factor in this study is the inclusion of brown coloured robots having ruled out social priming and task length as explanations for shooter bias, the implication is that the inclusion of another colour of robot reduced the shooter bias tendencies. It is particularly evident that this may be the case because a shooter bias with the Nao robot similar to Experiment A and previous research (Bartneck *et al.*, 2018) should have been expected, but there is no evidence of this here. I.e., even when considering the Nao robot separately there is no significant colour by object interaction for Nao robot ($F(2, 97) = 0.358, p = 0.700, \eta^2 = 0.007$).

Robot Race

After playing the game, participants were asked to ascribe a race to each of the three colours of Nao robot. Although the first option in the list was 'Does not apply', 75% of participants did ascribe a race to the robots. The robot racialised as 'White' was highly identified as White/Caucasian (66%), while the 'Black' robot was mainly ascribed to be Black/African American (27%) or Latino/Hispanic (22%). The light brown robot was mainly ascribed to be Latino/Hispanic (29%) with 11% ascribing it to be African American, and the rest ascribing it to one of the other races in the list that would normally be associated with a brown skin tone (e.g. Indian, Asian, Native American). Counts and percentages of responses are shown in table 5.3.

Table 5.3: Results of manipulation check for Robot Race - Experiment B

	Black		**White**		**Brown**	
Does not apply	*78*	25%	*65*	21%	*67*	21%
Caucasian/White	*11*	4%	*207*	66%	*3*	1%
Black/African	*85*	27%	*1*	0%	*35*	11%
Latino/Hispanic	*69*	22%	*7*	2%	*92*	29%
Native American	*17*	5%	*2*	1%	*21*	7%
Asian	*8*	3%	*27*	9%	*44*	14%
Indian	*27*	9%	*1*	0%	*31*	10%
Arab	*10*	3%	*1*	0%	*14*	4%
Pacific Islander	*7*	2%	*1*	0%	*5*	2%

It appears that introducing a brown coloured robot led to greater variability in the racialisation of both the Black and Brown coloured robots. Although participants clearly identified both colours as non-Caucasian, there was more variability across their racialisation as African American, Latino/Hispanic, or another ethnic/racial group.

Robot Anthropomorphism

A manipulation check was performed to investigate whether the robots used in the experiment were perceived differently according to their level of anthropomorphism. An ANOVA revealed no significant difference between the anthropomorphism ratings of the three robots, ($F(2, 622) = 2.183, p = 0.114, \eta^2 = 0.007$). A second ANOVA was performed in which only the ratings that each participant gave for the robot that they interacted with were compared. Here again, there was no significant effect of the robot type on the perceived anthropomorphism ($F(2, 309) = 2.163, p = 0.117, \eta^2 = 0.014$). Since the manipulation check revealed that the participants did not perceive the robots to be different in terms of their anthropomorphism, the perceived anthropomorphism was excluded from the further analysis.

However, although these different types of robot did not appear to vary on a continuum of anthropomorphism they may vary on an unmeasured construct. A $2 \times 3 \times 3$ ANOVA revealed a significant 3-way interaction between racialisation, object, and robot type on both reaction time ($F(4, 618) = 3.881, p = 0.004, \eta^2 = 0.025$) and accuracy ($F(4, 618) = 3.609, p = 0.006, \eta^2 = 0.023$) in the shooter bias paradigm. Decomposing this 3-way interaction, t-tests with Bonferroni corrected alpha revealed that there were only two significant simple effects, both toward Robosapien robots. Specifically, participants took significantly longer to refrain from shooting the Black Robosapien robot than the White ($p < .001$) or Brown Robosapien robots ($p < .001$). However, all other simple effects were non-significant across all three robot types.

Additionally, when testing whether there were differences in overall responses in the game toward the three types of robots it was found that robot type significantly impacted upon reaction time ($F(2, 309) = 6.3962, p = 0.002, \eta^2 = 0.040$) and accuracy ($F(2, 309) = 5.248, p = 0.006, \eta^2 = 0.033$) on the shooter bias task. Post-hoc t-tests with Bonferroni corrected alpha showed that participants took significantly longer to respond to the Nao robot than the Robosapien robot ($p = 0.001$), but they were significantly more accurate when responding to the Robosapien than the Inmoov robot ($p = 0.004$). Figure 5.1 shows how the average reaction times across all colours of robot show that participants were consistently faster to react to Robosapien, then Inmoov, then Nao whether they had a gun or not.

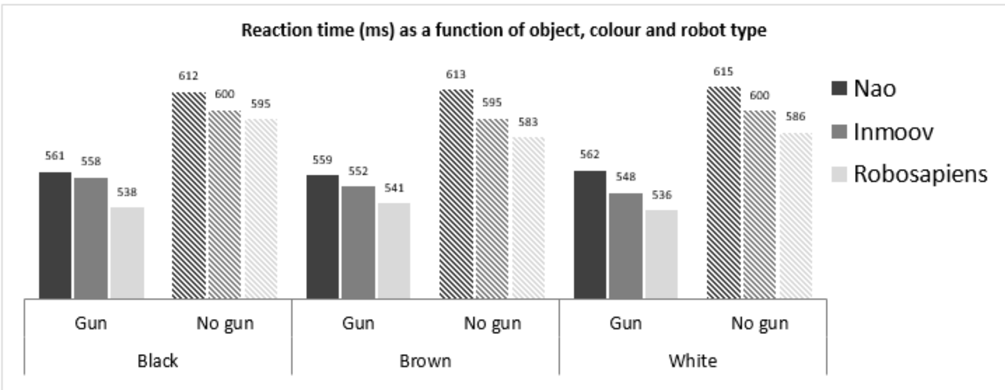


Figure 5.1: Average participant reaction times for the different robot types

5.7 Summary of Experiment B

To summarise, this experiment found that there was no shooter bias shown towards any particular robot type (Nao, Inmoov or Robosapien), and participants did not perceive them as more or less human-like than each other. However, there was a difference in reaction times between them. Also contrary to expectations, the addition of a third colour of robot seemed to remove the shooter bias completely.

Chapter 6: Conclusions, Limitations and Future Work

Experiment A sought to address issues in the preliminary robot racism research of possible social priming and whether player fatigue from task length had any influence on the shooter bias. The results showed that the shooter bias effect was still present for robots racialised as Black when there is no social priming, and that trial length does not have an influence on this race-based shooter bias. The experiment therefore succeeded in eliminating these as potential influencing factors and giving more validity to the previous results. Experiment B extended the research and sought to find out if the shooter bias varied across a racial spectrum as well as a range of anthropomorphism from less to more human-like.

Robot Anthropomorphism

Zlotowski *et al.* (2015) showed that humanoid robots were subject to the inversion effect, which measures how much people implicitly recognise objects as human. Given this, the idea that participants may be more hesitant to shoot a highly human-like robot compared to a machine-like robot was explored. A manipulation check was carried out with the expectation that participants would consciously view the three robots as differing in their perceived anthropomorphism, but this turned out not to be the case. There was also no shooter bias shown to any of the robot types, although there was a consistent difference in reaction times. They were faster to react to Robosapien and slower to react to Nao no matter which colour the robot was. This suggests that there may be a different dimension on which these robots implicitly varied, but future work is needed to more systematically examine the potential moderating role of robot anthropomorphism on shooter bias.

Does Diversity Reduce Bias?

There is a relationship between people's interracial exposure in their local environment to their underlying perceptions of race. The less interracial exposure people have, the more marked shifting between categories they displayed when asked to attribute race to mixed-race faces (Freeman & Sanchez, 2016). Gaertner & Dovidio (2005) found that children who had been through a program that helps bring people from different groups conceptually into their own circle of existence were more willing to pick playmates who were different than themselves. Researchers in China found that after training pre-school children how to individuate African faces, their implicit racial bias towards this out-group was significantly reduced (Xiao *et al.*, 2015).

Because of prior work that showed the shooter bias towards Black men was still present in a multi-ethnic context (Sadler *et al.*, 2012), the expectation was that the same would be found in the robot studies. However, the shooter bias towards the robots racialised as Black disappeared when a brown robot was present for all robot types. Additionally, participants were neither faster to shoot armed robots that were darker in colour, nor to not shoot unarmed robots that were lighter in colour. As the main variable that had not been excluded as an influence was the addition of the brown robot, this potentially means that diversification of robot colours may lead to a reduction in implicit racial bias towards them. However, future work is needed to further examine if the inclusion of a diverse range of colours on robots can indeed erase any colour-based bias that emerges from robot racialisation. Prior social-scientific research in human-human contexts suggest that increased exposure to diversity may have both positive and negative implications for implicit and explicit biases.

Galinsky *et al.* (2015) found that although diversity in race, culture and gender had potential social benefits, it also had negative effects such as causing conflict and resentment which could overshadow the positive impacts. They looked at diversity within groups of people, and the attitudes of individuals acquired through their experiences with living or working with different kinds of people. Both situations were round to promote deeper levels of complex thinking and more effective responses to new challenges, but also created a level of distrust and conflict. Rae *et al.* (2015) found that in U.S. states that had a larger proportion of Black residents, there existed stronger implicit and explicit in-group bias among both White and Black respondents.

In the context of HRI, it would be particularly valuable for future work to directly test whether exposure to robots of a diverse range of colours and perceived racialisation can indeed lead people to show no differences in their implicit and explicit responses. The robot shooter bias study could be extended to robots with a range of non-human skin tones, e.g. true black and white, and primary colours such as blue, green, and red to see if there is any particular bias shown toward any of them. Interestingly, participants across several studies seem to racialise robots even when given the option to indicate that the robot does not have a race suggesting a tendency to racialise robots, but having robots in non-human colours (e.g., green or blue) should remove such tendencies.

One of the challenges of the present work was that the brown robot was ascribed a plurality of racial and ethnic categories making it difficult to assess if participants' responses were due to stereotypes and prejudice toward one group or another. Indeed, these could lead to contradictory implicit biases making it important for future work to disentangle shooter biases toward different groups separately (i.e., among those that perceived the brown robot to be Latino vs. Indian vs. Native American). Future studies could also examine automatic perception of race in robots rather than rely on explicit means of race ascription. For example, the methods outlined in Freeman *et al.* (2010) in which mouse-tracking was used to measure how much subjects skewed towards one image or the other when asked to ascribed a race to human faces across a racial spectrum, could be one technique that allows researchers to better understand racialisation in the context of robotics. Also, as the present research uses online convenience samples, future work should examine these effects in a more representative national sample. The original research by Correll *et al.* (2002) showed that both African American and White American participants demonstrate comparable levels of shooter bias towards Black men, so a more ethnically diverse sample may not show differing effects. However, a more representative sample would allow researchers to better understand whether certain characteristics moderate such effects.

The shooter bias measured people's response to a perceived threat. While this worked well to heighten people's implicit reactions, a future experiment could look at the effect of working with different coloured robots. If a racism Implicit Association Test (IAT) was taken before and after solving a problem or completing a task with a racialised robot, any change in their subconscious racial attitudes could be measured. As found in the shooter bias studies, any initial results would need to be assessed for priming and a range of non-human colours should also be used as a control.

Final Thoughts

If we do perceive humanoid robots as racialised even subtly, it stands to reason that exposing people, especially children, to them in a variety of skin tones will help them subconsciously normalise different races. If social robots are going to be used widely, they should express the traits that best match the situation and purpose. Even if these traits are initially based on stereotypes, it may also help move us away from perceiving particular roles as gender or race based. So rather than replacing workers, social robots may help to dispel myths and stereotypes around societal roles and help combat bias.

If artificial intelligence is the way of the future, it should be leading the way in diversity and inclusiveness.

Appendix A: Forms and Documents

A.1 Participant Information and Consent Sheet

INFORMATION SHEET

ATTITUDES TOWARD TECHNOLOGY IN EVERYDAY LIFE

You are invited to participate in a research study conducted by researchers at the University of Canterbury. Please read the information below, which outlines what is involved in this research. If you would like to take part in this study, which will take approximately 20 minutes, give your consent by ticking the corresponding box on the Amazon Mechanical Turk form.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch; email human-ethics@canterbury.ac.nz. Any inquiries or complaints can be addressed to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch 8140.

PURPOSE OF THE STUDY

The purpose of this project is to examine attitudes toward the use of modern technology in everyday life.

PROCEDURE

If you choose to participate in this study, you will first be asked to complete a demographic survey. After this, you will complete a rapid judgment task where you will need to quickly determine whether an object is dangerous or not. And finally, you will complete a short questionnaire where you will be asked to give your impressions about this technology. The entire study will take 18-20 minutes to complete and will only involve this single session.

POTENTIAL RISKS AND DISCOMFORTS

There are no known risks associated with this research. Participation in this study is voluntary and your responses will be entirely confidential. In other words, your identity will never be revealed and your data will be reported in a manner that makes it impossible for others to identify your responses.

POTENTIAL BENEFITS TO PARTICPANTS AND ORGANISATIONS

The results of the study will be used to better understand people's feelings toward technology in everyday life. Additionally, you will receive $1.25 USD through your Amazon Mechanical Turk account upon full completion of the study.

CONFIDENTIALITY

The researchers are very mindful of the need to protect participants' interests. Any information that you provide will be completely anonymous and you will not be asked to indicate your name or address in any section of the study. Only the principal researchers and co-investigators will have access to the raw data. Under no circumstances will any data you supply be disclosed to a third party in any way that could reveal who the source was. Because this research involves anonymous questionnaires you can be assured that your personal information will not be revealed in any report.

PARTICIPATION AND WITHDRAWAL

Participation is entirely voluntary. If you volunteer to be in this study, you may withdraw at any time without consequence of any kind. If you wish to withdraw from the study and have your responses deleted, simply exit the browser you are working on. If you decide that you wish to have your data deleted at the end of the study, you can indicate so on the feedback box on the Mechanical Turk form.

Figure A.1: Information

Consent

CONSENT TO PARTICIPATE IN ATTITUDES TOWARD TECHNOLOGY IN EVERYDAY LIFE

- I have read and understood the description of the above-mentioned project.
- I understand that my participation will involve completing an anonymous questionnaire.
- I understand that I will receive $1.25 USD upon full completion of the study.
- I understand that I can withdraw from the study at any time and the data I provided will be deleted.
- I agree to publication of results, with the understanding that my anonymity will be preserved.
- I also understand and am satisfied with all the measures that will be taken to protect my identity and ensure that my interests are protected.

I fully accept that I am giving my consent to participate in this research study. By ticking the corresponding box on the Amazon Mechanical Turk form, I indicate that I understand and agree to the conditions listed above.

Figure A.2: Consent

A.2 Participant Debriefing Sheet

Debriefing

You have just participated in a study examining people's general attitudes towards robots. Though we were interested in the same topic, we had other interests that we were unable to tell you about until now. Specifically, the aim of this study was to examine whether manipulating the characteristics of the robot affected people's feelings and attitudes toward robots and its perceived danger.

The reason we withheld this information from you is because: firstly, we examined whether making a robot more human-like would decrease negative attitudes and feelings towards robots. As such, we needed to show some participants more humanoid robots and others more machine-like robots. Secondly, we examined whether the colour of the robot additionally influenced attitudes toward the robot and rapid judgments of its perceived danger. In order to ensure that you were not unconsciously influenced to do this, we withheld this information from you.

You may be curious about our hypotheses. Research shows that during human robot interactions, people feel more comfortable and more trusting in the presence of a human-like robot than a machine-like robot. These feelings, in turn, facilitate human robot interaction. Additionally, other research reveals that people have an implicit or unconscious preference for lighter skin over darker skin evident in rapid reaction time tasks. Here, we wanted to examine how these findings map onto attitudes and rapid judgments of inanimate beings such as robots. Therefore, we wanted to test the hypothesis that darker coloured robots will be more negatively evaluated than lighter coloured robots, but only when the robot appears more human-like, and not when it is more machine-like. We would like to point out that the robots you saw did not have any special abilities and do not pose any threat to you or society.

Figure A.3: Debriefing

A.3 Publication and Contribution Details

ACM Proceedings (acmpubs MB) - Submission Follow Up, ID No: aies644

Actions

Mik Brunstetter [submissions@scomminc.c...

To: arifah.addison@hotmail.com

Cc: Arifah Addison; Christoph Bartneck; Kumar Yogeeswaran

Attachments: aies644-addisonA.pdf (617 KB) [Open as Web Page]

Friday, 5 April 2019 9:15 a.m.

Dear Author(s),

Thank you for the submission of "Robots Can Be More Than Black And White: Examining Racial Bias Towards Robots" to the proceedings and ACM Digital Library.

We received notice of your completed ACM form, so everything appears to be in order with the pdf submitted (attached) and it is being utilized for the conference publication(s) and ACM DL. No revisions are needed, nor will be accepted. This submission is being moved onto the next stages of production (but kindly note the signed ACM forms still go through a double-check by ACM for various permissions).

Thank you,

Mik Brunstetter

Sheridan Communications

Figure A.4: ACM Confirmation of Publication

Contributions to Publication (in POS format as applicable)

Arifah Addison - Student	AA
Christoph Bartneck - Senior supervisor (HITLabNZ)	CB
Kumar Yogeeswaran - Co-supervisor (Psychology Dpt)	KY

Conceptualisation	Ideas; formulation or evolution of overarching research goals and aims.	AA, CB/KY	Initial idea of study was outlined by the supervisors prior to project as a followup to their previous paper. This was picked by student and fleshed out together with subsequent goals evolving along the way based on initial results and analysis.
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse.	AA	All done by student. Data transformation process (database and queries) created for first experiment, then subsequently replicated for each new one.
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.	AA, KY/CB	Basic data analysis and inspection done by student, with final statistical analysis driven by supervisors initially, student performing subsequently after learning the process.
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.	AA	All done by student
Methodology	Development or design of methodology; creation of models	AA	All done by student after discussions with supervisors.
Project Administration	Management and coordination responsibility for the research activity planning and execution.	AA	All done by student
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.	AA	All done by student. Code for main experiments adapted from previous study. New code developed for supporting areas (e.g HTML for web pages and SQL queries for data transformation).
Visualisation	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.	AA	All done by student, with revision from feedback by supervisors.
Writing – Original Draft Preparation	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).	AA, CB/KY	All contributed to paper for publication as applicable. Supervisors ensured accuracy of interpretation of data and information. Student responsible for submitting final draft and making edits after feedback. Adaptation of paper into thesis done by AA, including adding background, detailed methodology and literature reviews, and editing and refining after feedback from supervisors.
Writing – Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.		

Figure A.5: Co-Author Contributions

Appendix B: Screen Shots

B.1 Game-play Instructions for Experiment A

Apart from the main instructions at the beginning of the experiment, participants also received additional messages between each phase of the robot shooter game to ensure a smooth and informed process. The full set of instructions and messages encountered in Experiment A which had two critical rounds are shown here.

PRACTICE ROUND.

Reminder:

You need to decide as quickly and accurately as possible to shoot or not to shoot.

- If the robot is holding a GUN you need to SHOOT: press the "A" key on your keyboard with your LEFT hand.
- If the robot is holding a HARMLESS OBJECT, do NOT shoot: press the "L" key on your keyboard with your RIGHT hand.

IMPORTANT: use fingers of different hands for each response key.

Figure B.1: Practice Round

Great, you have finished the practice round and are ready for the critical rounds.

Reminder:

You need to decide as quickly and accurately as possible to shoot or not to shoot.

- If the robot is holding a GUN you need to SHOOT (press the "A" key on your keyboard with your LEFT hand).
- If the robot is holding a HARMLESS OBJECT, do NOT shoot (press the "L" key on your keyboard with your RIGHT hand).

IMPORTANT: use fingers of different hands for each response key.

If you are ready, press the space bar to begin.

Figure B.2: Critical Round One

Before you proceed to the next round, take some time to rest.

The tasks you are completing are designed to be difficult and no one is expected to perform it perfectly, so just relax and do your best.

When you are ready, press the spacebar to continue to the next round.

Figure B.3: Between Critical Rounds

ROUND 2 / 2

Reminder:
You need to decide as quickly and accurately as possible to shoot or not to shoot.

- If the robot is holding a GUN you need to SHOOT: press the "A" key on your keyboard.

- If the robot is holding a HARMLESS OBJECT, do NOT shoot: press the "L" key on your keyboard.

IMPORTANT: use fingers of different hands for each response key.

If you are ready, push the space bar to begin.

Figure B.4: Critical Round Two

Great, you have finished the game!

Your final score was -20 points.

We now want to get your impression of some different robots. You will see a series of images and some corresponding questions.

All questions are required so you must choose an option for every question in order to move to the next image (so even if a slider is at the setting you want you must move it away and back again).

Remember that the study is anonymous and there are no right or wrong answers, we just want to get your opinions.

Press the space bar to see the first image

Figure B.5: End of Game

B.2 Inquisit Files

All the files loaded needed for the Inquisit experiment.

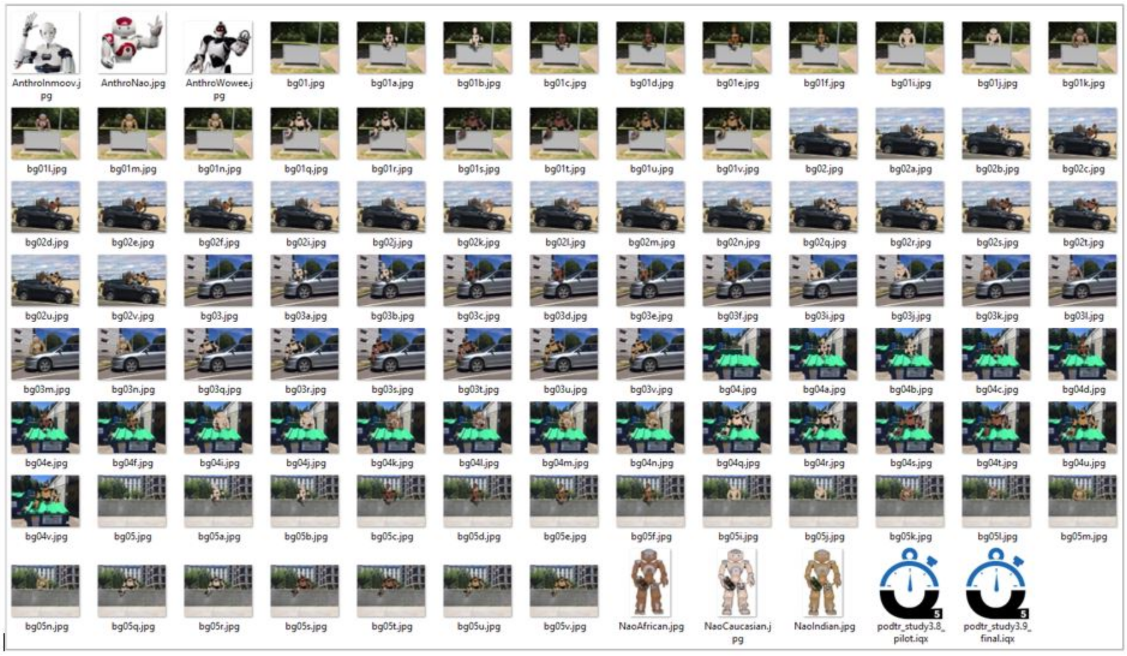


Figure B.6: Files for Experiment B

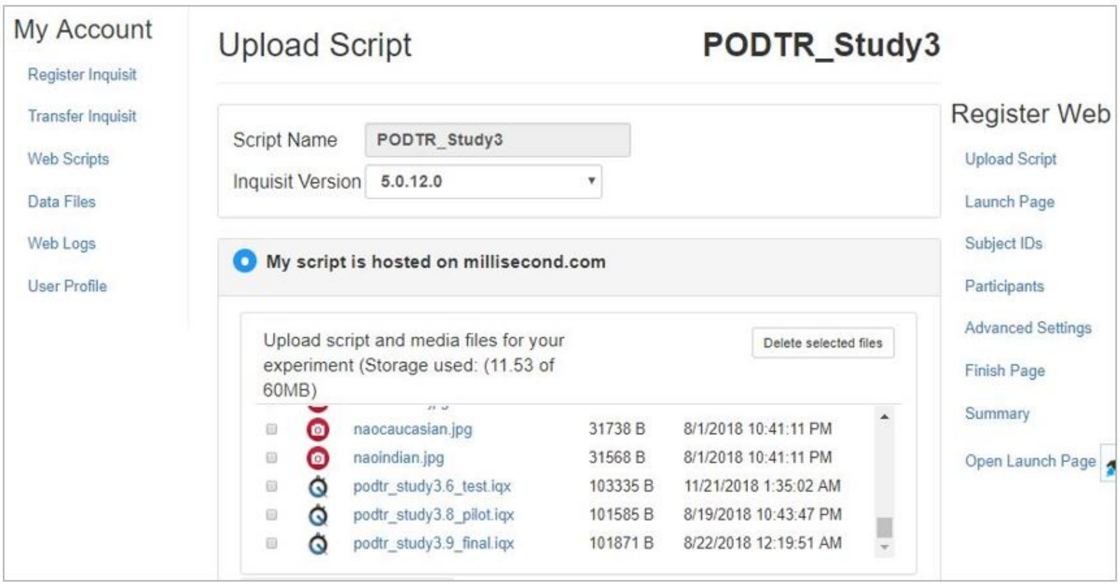


Figure B.7: Uploaded files in Inquisit Web

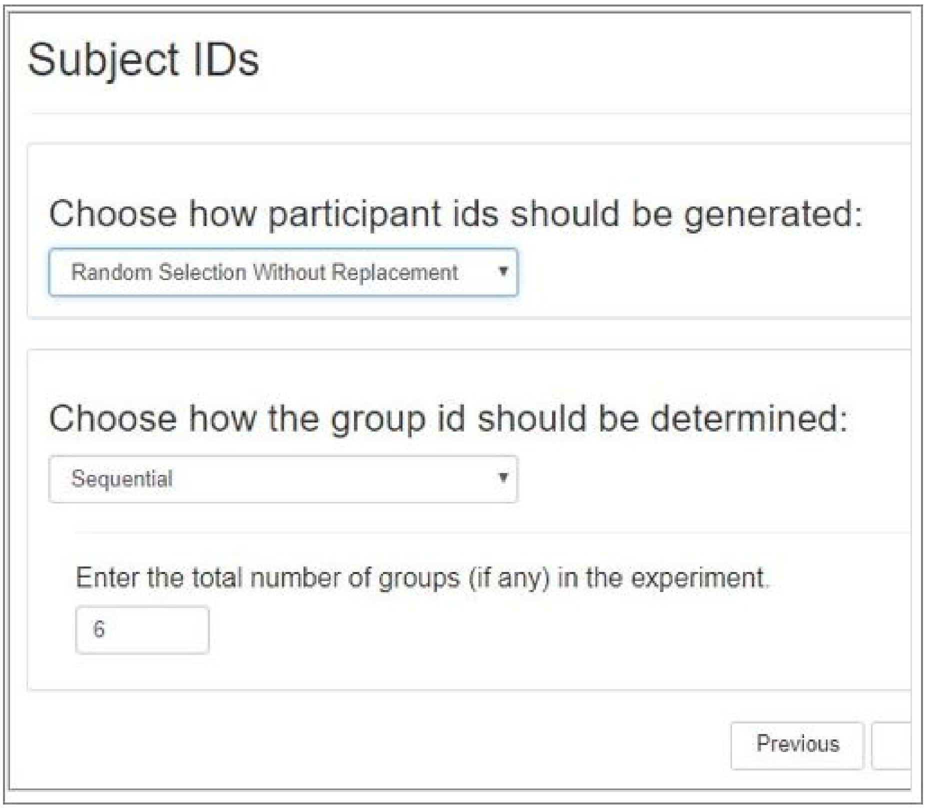


Figure B.8: Inquisit Web Subject ID settings for Experiment B

Appendix C: Data Dictionaries

C.1 Inquisit Data Files

Each field in the output data files generated by the Inquisit task for Experiment B are explained below. There are fields for general data and the three phases of the experiment - the demographics survey, the robot shooter game-play, and manipulation checks for perceived robot race and anthropomorphism.

General data: Computer and date/time information as well as the Subect ID and generated by the experiment, and their unique Inquisit code that is calculated from this using a user-defined formula.

Demographics: MturkID (Mechanical Turk WorkerID for cross-referencing), Gender, Age, Handedness (left or right), Race, Country, Nationality, Religion.

Game-play: The robot shooter game stimulus and responses, what key they hit and whether it was correct, the result code, and the latency in milliseconds.

Manipulation Checks: Robot Race (3 colours of robot, light - dark), Robot Anthropomorphism (3 types of robot) - 5 godspeed questions for human-ness + danger and trustworthiness.

Data File Information

**fields named 'values.xxx' are user-defined (as opposed to automatic system fields).*

(1) Summary data file: 'podtr_summary*.iqdat' (a separate file for each participant).

values.uniquecode:	Allocated unique code for verification (formula in script)
script.subjectid, script.groupid:	Auto allocated subjectid (random) and groupid (1-6)
script.startdate:	Date script was run
script.starttime:	Time script was started
script.elapsedtime:	Time it took to run script (in ms)
computer.platform:	Platform the script was run on
/completed:	Was the game completed or not
/total:	Total points (final total)
The following	
/countGun** - /countObject**:	Counts the number of trials for each robot colour/type
/Sum**Hits - /Sum**CRs:	Running total of Hits, Misses, FAs, CRs for each robot colour/type
/MeanRT_**Hits - /MeanRT_**CRs:	Mean reaction times for Hits, Misses, FAs, CRs for each robot colour/type (in ms)
/percent_**Hits- /percent_**CRs:	Percentages of Hits-Misses and CRs-FAs, separately for each robot colour/type

(2) Raw data file: 'podtr_raw*.iqdat' (a separate file for each participant).

Field	Description	Example data
build	Inquisit build	**5.0.12.0**
computer.platform	Platform script was run on	**Mac, Win**
date	Date script was run	**080618** (MMDDYY) **leading zero is dropped*
time	Time script was run	**09:26:34** **stored as text when downloaded to Excel*
values.uniquecode	Allocated unique code for verification in MTurk	**POD42856247302** **formula is defined in the code*
script.subjectid	System-allocated SubjectID	**952081735**
script.groupid:	System-allocated GroupID (1-6)	**1-6** as defined in the settings for the Inquisit Web Script **6 corresponding EXP blocks in the code, 2 for each robot to counterbalance for L/R shootkey*
values.group values.robot	Robot group number (1-3) and Robot type (text)	**1 - Inmoov, 2 - Nao, 3 – Wowee** As defined in the corresponding EXP blocks
blockcode blocknum	Name of Block Block number	As defined in the EXPs: / blocks = [**1=Demographics; 2=POD_practice; 3=POD_1; 4=POD_2; 5=ManipulationBlockAnthro; 6=ManipulationBlockRace**]
trialcode trialnum	Name and number of the currently recorded trial	Specific data being recorded, can be gameplay (eg **4 - ObjectBrownNao** or Survey Questions (eg **1 - gender**)
values.itemnumber	Itemnumber of the target image (gameplay trials only)	Depends on number of items listed **1-20** for practice, **1-10** for critical
values.selectbackground	Item number of background picture selected	Depends on number of items listed **1-20** for practice, **1-10** for critical **only differs from **itemnumber** for practice rounds, is identical for critical rounds*
values.repetition	Number of empty background images presented prior to robot stimuls image.	1-4 (random, as defined in the code)
values.background duration	How long each background slide was presented	500-1000 (radom, increments of 100)
values.count_background	How many background slides were presented	1-4 (defined in code) 1-20 for practice, 1-10 for critical **only differs from **repetition** for practice rounds, is identical for critical rounds*

response	**The participant's response**	Actual data recorded **Demographics**: Whatever they entered for each survey question (age, gender etc). **Anthro**: radio button or slider choices (eg 1-5, 1-10 depending on scale) **Gameplay**: Keypress# – should be **30** or **38** for "A" and "L" keys, or **0** for nothing pressed (too slow).
correct	The correctness of the response (redundant for survey questions)	(1 = correct; 0 = incorrect) Gameplay – unclear what makes it 'correct'
values.result	Gameplay blocks only, result of current response	If 'correct'=0: "**Miss**" (did not shoot a bad guy), "**FA**" (shot a good guy), "**NoResponse**" If 'correct'=1: "**Hit**" (shot a bad guy) or "**CR**" (did not shoot a good guy)

The 'result' can be one of 5, corresponding with the Feedback item list, and correllating with value of 'correct'

Incorrect responses (correct=0)

- **NoResponse** = did not hit either key in time ("Too Slow")
- **Miss** = missed a bad guy: pressed 'not shoot' key for robot with gun ("You're Dead!!!")
- **FA** (False Alarm) = shot a good guy: pressed 'shoot' key for robot without gun ("You shot a good guy!")

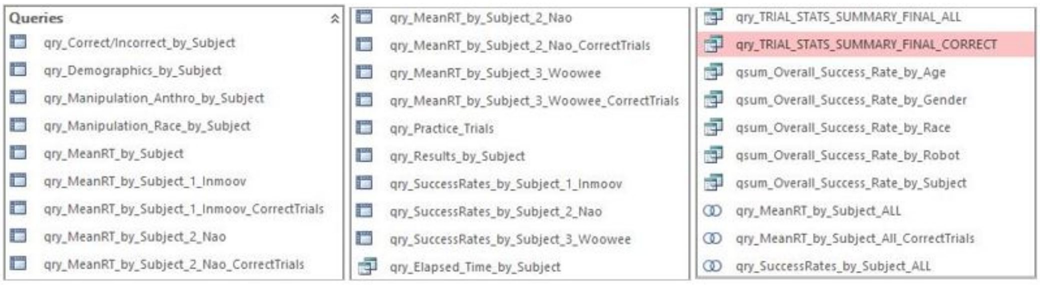
Correct responses (correct=1)

- **Hit** = shot the bad guy: pressed 'shoot' key for robot with guy ("Good Shot")
- **CR** (Correct Rejection) = did not shoot a good guy: pressed 'not shoot' key for robot without gun ("Wise Choice")

latency:	The response latency (in ms)	Redundant for survey questions
values.total	Running total as of current trial	-1000 – 1000 (grows as game progresses)

C.2 Output Data Files

Query List Various queries extract data for each block and robot type as well as overall summaries. The queries used in Experiment B are shown below.



The final output query with the transformed data fields are listed below, with example values and descriptions of what they are made up or and/or used for.

Output Field	Example Value	Description
SubjectID	14477530	Unique SubjectID generated by Inquisit
WorkerId	A38IB23GPSEKY1	Unique Mturk WorkerID entered by the participant
InquisitCode	POD243434687	Unique ID generated within the game code using the following formula: "POD GroupID SubjectID x factor = + factor = InquisitCode *Eg:* POD 3 344032169 3 1032096507 2097 1032098604 POD31032098604
Age	31	Demographics responses
Gender	Male	
Gender_Code	1	Gender must be coded (Male=1, Female=0) so it can be used in analysis
Hand	Right	
Nationality	American	
Race	White/Caucasian	
Religion	christian	
State	michigan	
ScriptGroup	2	Script Group number 1-6 (sequential)
Robot_Group	1	Robot group, one of the 3 types (random)
Robot	Inmoov	Name of robot corresponding to Robot Group number (number is needed for the analysis)
Score	450	Final total score
Noresponse	0	Number of 'no response' gameplay results (i.e they didn't hit any key in time).
Num_Correct	60	Number of correct gameplay responses
Overall Success Rate	100%	Overall success rate (Num_Correct / total possible game-play responses)
Mean Reaction Time	490.1333333	Average reaction time across all correct responses
GunBlack_SR_mean	100%	Success rates for each condition
ObjectBlack_SR_mean	100%	
GunBrown_SR_mean	100%	
ObjectBrown_SR_mean	100%	
GunWhite_SR_mean	100%	
ObjectWhite_SR_mean	100%	
GunBlack_RT_mean	444.7	Average reaction times for each condtion
ObjectBlack_RT_mean	520.3	
GunBrown_RT_mean	491.5	
ObjectBrown_RT_mean	495.3	
GunWhite_RT_mean	464.8	
ObjectWhite_RT_mean	524.2	
Log_GunBlack_RT	6.097399898	Log-transformed average reaction times for each condition
Log_ObjBlack_RT	6.254405568	
Log_GunBrown_RT	6.19746194	
Log_ObjBrown_RT	6.20516364	
Log_GunWhite_RT	6.141607206	
Log_ObjWhite_RT	6.261873291	

Figure C.1: Data output for analysis - main measures

RobotRace1_Black	Pacific Islander	Manipulation checks for perception of Robot Race (1 for each colour)
RobotRace2_White	Caucasian/White	
RobotRace3_Brown	Latino/Hispanic	
Own_AnthroQ1_Natural	1	Manipulation checks for perception of Robot Anthropomorphism: - responses to their 'own' robot (the one they encountered in game-play) - responses to each type of robot (1-3) - average for each - percieved traits (danger/trust) for each 1-5 for anthroporphism, 1-10 for danger/trust, as shown.
Own_AnthroQ2_Humanlike	1	
Own_AnthroQ3_Conscious	2	
Own_AnthroQ4_Lifelike	1	
Own_AnthroQ5_Elegant	2	
Anthro_Own	7	
Own_Anthro_Danger	4	
Own_Anthro_Trust	5	
1_AnthroQ1_Natural	1	
1_AnthroQ2_Humanlike	1	
1_AnthroQ3_Conscious	2	
1_AnthroQ4_Lifelike	1	
1_AnthroQ5_Elegant	2	
1_Anthro_Inmoov	7	
1_Anthro_Danger	4	
1_Anthro_Trust	5	
2_AnthroQ1_Natural	1	
2_AnthroQ2_Humanlike	2	
2_AnthroQ3_Conscious	1	
2_AnthroQ4_Lifelike	1	
2_AnthroQ5_Elegant	2	
2_Anthro_Nao	7	
2_Anthro_Danger	1	
2_Anthro_Trust	6	
3_AnthroQ1_Natural	1	
3_AnthroQ2_Humanlike	2	
3_AnthroQ3_Conscious	2	
3_AnthroQ4_Lifelike	2	
3_AnthroQ5_Elegant	3	
3_Anthro_Wowee	10	
3_Anthro_Danger	7	
3_Anthro_Trust	5	

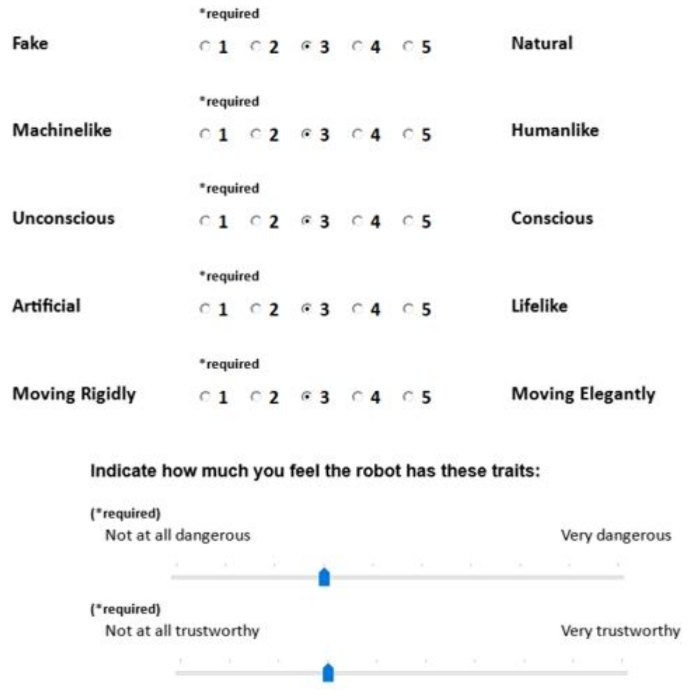


Figure C.2: Data output for analysis - manipulation checks

References

Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, **1**, 71–81. 18, 31

Bartneck, C., Duenser, A., Moltchanova, E. & Zawieska, K. (2015). Comparing the similarity of responses received from experiments conducted in amazon's mechanical turk to experiments conducted with traditional methods. *PLOS One*, **10**, e0121595.

Bartneck, C., Yogeeswaran, K., Sparrow, R., Wang, S. & Eyssel, F. (2018). Robots and racism. In *ACM/IEEE International Conference on Human-Robot Interaction*, 196–204. i, 2, 4, 6, 7, 28, 33

Cabibihan, J.J., Javed, H., Ang, M. & Aljunied, S.M. (2013). Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics*, **5**, 593–618. 3

Cameron, W.B. (1963). *Informal sociology, a casual introduction to sociological thinking*. Random House studies in sociology, Random House, New York,.

Converse, J.M. & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*, vol. 63. Sage, Thousand Oaks, California.

Correll, J., Park, B., Judd, C.M. & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, **83**, 1314–1329. 2, 6, 27, 28, 36

Correll, J., Urland, G.R. & Ito, T.A. (2006). Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, **42**, 120–128.

Correll, J., Ma, D.S., Wittenbrink, B., Bar-Anan, Y. & Nosek, B.A. (2013). When fatigue turns deadly: The association between fatigue and racial bias in the decision to shoot. *Basic and Applied Psychology*, **35**, 515–524. 5

Derman Sparks, L., Higa, C. & Sparks, B. (1978). Children, race and racism: How race awareness develops.

Eyssel, F. & Kuchenbrandt, D. (2011). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, **51**, 724731. 3

Freeman, J.B. & Johnson, K.L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences*, **20**, 362–374.

Freeman, J.B., Pauker, K., Apfelbaum, E.P. & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, **46**, 179–185. 36

Freeman, P.K., Jonathan B. & Sanchez, D.T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science*, **27**, 502–517. 35

Gaertner, S.L. & Dovidio, J.F. (2005). Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model. *Journal of Social Issues*, **61**, 615–639. 2, 35

Galinsky, A.D., Todd, A.R., Homan, A.C., Phillips, K.W., Apfelbaum, E.P., Sasaki, S.J., Richeson, J.A., Olayon, J.B. & Maddux, W.W. (2015). Maximizing the gains and minimizing the pains of diversity a policy perspective. *Perspectives on Psychological Science*, **10**, 742748. 36

Gerrold, D. (2018). My friend, the robot. 1

Gilbert, D.T. & Hixon, G.J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, **60**, 509–517. 5

Gong, L. (2008). The boundary of racial prejudice - comparing preferences for white, black and robot characters. *Science Direct*, **24**, 20742093. 1

Gordon, G., Spaulding, S., Westlund, J.K., Lee, J., Plummer, L., Martinez, M. & Breazeal, C. (2016). Affective personalization of a social robot tutor for childrens second language skills. In *Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, 3951–3957, AAAI Press, Phoenix. 3

Hewstone, M. (2015). Consequences of diversity for social cohesion and prejudice: The missing dimension of intergroup contact. *Journal of Social Issues*, **71**, 417–438.

Lai, C., Marini, M., Lehr, S., Cerruti, C., Shin, J.E., Joy-Gaba, J.A. & Nosek, B. (2014). Reducing implicit racial preferences: I. a comparative investigation of 17 interventions. *Journal of Experimental Social Psychology: General*, **143**, 1765–1785.

Lazerus, T., Ingbretsen, Z.A., Stolier, R.M., Freeman, J.B. & Cikara, M. (2016). Positivity bias in judging ingroup members emotional expressions. *Emotion*, **16**, 1117–1125.

Macrae, N.C., Milne, A.B. & Bodenhausen, G.V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, **66**, 37–47. 5

Mavridis, N., AlDhaheri, A., AlDhaheri, L., Khanii, M. & AlDarmaki, N. (2011). Transforming ibnsina into an advanced multilingual interactive android robot. In *2011 IEEE GCC Conference and Exhibition (GCC)*, 120–123, IEEE, New York, NY, USA.

Mekawi, Y. & Bresin, K. (2015). Is the evidence from racial bias shooting task studies a smoking gun? results from a meta-analysis. *Journal of Experimental Social Psychology*, **61**, 120–130. 2

Ogunyale, B.D., T. & Howard, A. (2018). Does removing stereotype priming remove bias? a pilot human-robot interaction study. *ArXiv e-prints*. 4, 26

Rae, J.R., Newheiser, A.K., & Olson, K.R. (2015). Exposure to racial out-groups and implicit race bias in the united states. *Social Psychological and Personality Science*, **6**, 535–543. 36

Sadler, M.S., Correll, J., Park, B. & Judd, C.M. (2012). The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues*, **68**, 286–313. 4, 7, 30, 35

Sánchez Ramos, A.C., Contreras, V., Santos, A., Aguillon, C., Garcia, N., Rodriguez, J.D., Amaya Vazquez, I. & Strait, M.K. (2018). A preliminary study of the effects of racialization and humanness on the verbal abuse of female-gendered robots. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*, 227–228, ACM, New York, NY, USA.

Strait, M., Sanchez Ramos, A., Contreras, V. & Garcia, N. (2018). Robots racialized in the likeness of marginalized social identities are subject to greater dehumanization than those racialized as white. In *IEEE International Conference on Robot and Human Interactive Communication*. 1

Tadmor, C., Hong, Y.y., Chao, W.F., M.M. & Wang, W. (2012). Multicultural experiences reduce intergroup bias through epistemic unfreezing. *Journal of Personality and Social Psychology*, **103**, 750–772.

Tapus, A., Mataric, J. & Scassellati, B. (2007). The grand challenges in socially assistive robotics. *IEEE Robotics and Automation Magazine*, **14**, 35–42. 3

Tay, B., Jung, Y. & Park, T. (2014). When stereotypes meet robots - the double-edged sword of robot gender and personality in human-robot interaction. *Computers in Human Behaviour*, **38**, 75–84. 3

Trovato, G. & Eyssel, F. (2017). Mind attribution to androids: A comparative study with italian and japanese adolescents. In *26th IEEE International Symposium on Robot and Human Interactive Communication*, 561–566.

Unkelbach, C., Forgas, J.P. & Denson, F., Thomas (2008). The turban effect: The influence of muslim headgear and induced affect on aggressive responses in the shooter bias paradigm. *Journal of Experimental Social Psychology*, **44**, 1409–1413.

von der Pütten, A.M. & Krämer, N.C. (2012). A survey on robot appearances. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, 267–268, ACM, New York, NY, USA.

Waytz, A., Cacioppo, J. & Epley, N. (2014). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives in Psychological Science*, **5**, 219–232.

Weir, K. (2018). The dawn of social robots. *Monitor on Psychology*, **49**. 1

Whitaker, R.M., Colombo, G.B. & Rand, D.G. (2018). Indirect reciprocity and the evolution of prejudicial groups. *Scientific Reports*, **8**, 13247.

Xiao, W.S., Fu, G., Quinn, P.C. & Qin, J. (2015). Individuation training with other-race faces reduces preschoolers implicit racial bias: a link between perceptual and social representation of faces in children. *Developmental Science*, **18**, 655–663. 35

Yogeeswaran, K., Devos, T. & Nash, K. (2017). *Understanding the Nature, Measurement, and Utility of Implicit Intergroup Biases*. Cambridge University Press. 2

Zlotowski, J. & Bartneck, C. (2013). The inversion effect in hri: Are robots perceived more like humans or objects? In *ACM/IEEE International Conference on Human-Robot Interaction*, 365–372. 4

Zlotowski, J., Bartneck, C., Yogeeswaran, K. & Proudfoot, D. (2015). Anthropomorphism: Opportunities and challenges in human-robot interaction. *International journal of social robotics*, **7**, 347–360. 35